E. Galéev, V. Tijomírov

Breve curso de la TEORIA de PROBLEMAS EXTREMALES

Editorial Mir Moscú

Breve curso de la teoría de problemas extremales

Э. М. Галеев, В. М. Тихомиров

КРАТКИЙ КУРС ТЕОРИИ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

Издательство Московского университета

E. Galéev, V. Tijomírov

Breve curso de la teoría de problemas extremales

Editorial Mir
Moscú

Traducido del ruso por
A. Cherkásov

Impreso en la URSS

На испанском языке

ISBN 5-03-001957-X (исп.)

ISBN 5-211-00313-6 (русс.)

INDICE

PREFACIO

Tras haber adquirido enorme importancia en la actualidad, los problemas de optimización pasaron a ser un factor que predeterminó la necesidad imprescindible de recurrir a elementos de los problemas extremales en la instrucción matemática, de ingeniería y económica.

La finalidad del presente libro consiste en contribuir a que la citada teoría ocupe su digno lugar en la instrucción matemática actual.

La primera parte engloba las principales ramas de la teoría que entran comúnmente en los programas de los cursos de análisis matemático y de diferentes cursos de optimización: la regla de factores de Lagrange para problemas con igualdades y desigualdades (§ 2), la programación lineal (§ 3), el cálculo de variaciones clásico (§§ 4, 5) y el control óptimo (§ 6). Esa parte, especialmente sus primeros cuatro párrafos, está destinada al más amplio auditorio Quisiéramos que el presente material se usara para estructurar tanto los cursos de optimización como los fragmentos independientes que se incluyan en los estudios de matemática generales de los centros de enseñanza superior técnicos y económicos. La asimilación del material de la primera parte da la posibilidad de resolver tales problemas.

Cabe señalar que el material de la primera parte agota prácticamente el programa del curso de «Cálculo de variaciones y métodos de optimización» de la especialidad 2013 de «Matemáticas» para las universidades estatales.

La segunda parte (§§ 7—11) del manual está destinada, ante todo, a los estudiantes y el personal docente encargado de los cursos de optimización en las universidades y otros centros de enseñanza superior. Esa parte puede aplicarse estructurando tanto los cursos básicos como los estudios especiales relacionados con la referida optimización. Uno de los objetivos de la instrucción universitaria consiste en descubrir las causas profundas de una u otra teoría y el lugar que esta última ocupa en la estructura de toda la ciencia matemática. En la segunda parte se intenta de hacer todo eso en relación con la teoría de problemas extremales, sometiendo aquí a un nuevo enfoque el material de la primera parte desde el único punto de vista que reúne el análisis plano y convexo infinitésimo. Para llevar a cabo ejercicios prácticos y trabajos de laboratorio referentes a los métodos de optimización, se ofrece la solución de importantes problemas extremales (§ 12), además hay 260 problemas con sus resultados y con las soluciones de algunos de ellos.

En el manual se refleja la experiencia de impartir cursos de optimización en la facultad mecánico-matemática de la Universidad Estatal de Moscú. El autor de la primera parte, así como del § 12 y los problemas es E. M. Galéev Lo demás pertenece a la pluma de V M. Tijomírov

INTRODUCCION

1. Ensayo histórico. Los problemas de búsqueda de magnitudes máximas y mínimas fueron objeto de meticulosa consideración en el transcurso de toda la evolución de las matemáticas. Pero es especial la importancia que ellos han adquirido en nuestros días cuando se releva, como nunca antes, la necesidad en la más efectiva utilización de las riquezas naturales, los recursos humanos y los medios materiales y técnicos. Todo eso requiere dedicar más atención a los problemas de control. Y cada vez que es posible la participación activa del hombre, surge el deseo de buscar el mejor control o, según suele decirse, el control óptimo. En tales casos se recurre inevitablemente a las matemáticas.

El más antiguo de los problemas de máximo y mínimo se considera el llamado problema isoperimétrico clásico: hallar, entre las curvas planas cerradas de una longitud dada, la curva que abarca la mayor superficie. Los filósofos griegos trataron de resolver este problema aún en el siglo V a.n.e. Sobre él escribía el gran filósofo griego Aristóteles. Los geómetras griegos plantearon y resolvieron varios problemas extremales más, los cuales fueron reflejados en los «Elementos» de Euclides y en las obras de Arquímedes y Apolonio.

En la época de Renacimiento, tan pronto se renovaron las actividades científicas, muchos matemáticos se dedicaron de nuevo a los problemas de máximo y mínimo.

En el siglo XVII precisamente los problemas de ciencias naturales predeterminaron la necesidad de proseguir investigando los problemas extremales. Todo esto comenzó por los intentos de explicar la ley de refracción de la luz. Aún los antiguos trataron de deducirla. En particular, Ptolomeo (siglo II a.n.e.) intentó hallarla experimentalmente, pero su empresa fracasó. La ley de refracción de la luz fue establecida por el científico holandés Snellius en el siglo XVII y entonces surgió el problema de los principios físicos en los que ella se basa. Para explicar dicha ley, el gran matemático francés Pierre Fermat propuso (en el año 1600 aproximadamente) el principio extremal que recibió luego su nombre. El principio de Fermat dice: el trayecto que sigue un rayo luminoso en un medio heterogéneo es el más corto posible y corresponde a un tiempo mínimo.

Partiendo de ese momento, la idea de los llamados «principios variacionales» (es decir, la convicción de que las leyes de la naturaleza se deducen de correlaciones extremales) pasa a ser una de las ideas centrales en todas las ciencias naturales.

Hasta la segunda mitad del siglo XVII no hubo algún procedimiento general que permitiera solucionar los problemas extremales. La necesidad de formularlos estimuló notablemente la creación del análisis matemático.

El primer procedimiento de carácter general que se proponía para el estudio de los problemas de máximo y mínimo fue descrito por P. Fermat (en el año 1630). En el lenguaje actual su definición es la siguiente: en el punto extremal (de cierta función de una variable), la derivada es igual a cero y, por lo tanto, los extremos deben buscarse entre las raíces de las derivadas. Este resultado entra ahora en el curso escolar de matemáticas bajo el nombre de teorema de Fermat. Pero en realidad Fermat describió ese procedimiento tan sólo para los polinomios. En su forma general el mismo ha sido obtenido primeramente por Newton (en los años sesenta del siglo XVII), después fue descubierto de nuevo por Leibniz y publicado por él en el famoso artículo a partir del cual comienza la historia del análisis matemático. Es digno de atención el título de dicho artículo: «*Nova methodus promáximis et minimis*». En el siglo XVII Euler y Lagrange crearon los procedimientos de solución de problemas extremales con funciones de diversas variables sin acotaciones y con acotaciones del tipo de ecuaciones.

El principial procedimiento, denominado método de factores de Lagrange, ahora forma parte del programa de cualquier centro técnico o matemático de enseñanza superior. Ya en nuestros días tales estudios fueron completados por problemas donde las acotaciones se dan en forma de igualdades y desigualdades. Todo ese cielo de problemas es conocido con el nombre de programación matemática.

En la revista «*Acta Eruditorum*» del mes de junio de 1696 (primera y única revista científica de aquel entonces) apareció el artículo de Johann Bernoulli, conocido matemático, discípulo y continuador de Leibniz, bajo el siguiente título: «Un nuevo problema a cuya solución se invitan los matemáticos». Allí se planteaba el problema de la curva del descenso más rápido, o sea, el problema de la braquistócrona, a partir del cual comienza la historia del cálculo de variaciones clásico.

Los métodos generales de solución de problemas de cálculo de variaciones fueron elaborados en el siglo XVIII por Euler y Lagrange. Precisamente en esos trabajos fue establecida una estrecha relación entre el cálculo de variaciones y las ciencias naturales. La elaboración de la teoría de cálculo de variaciones prosiguió posteriormente en el transcurso de más de dos siglos. Además de las condiciones necesarias de primer orden (ecuaciones de Euler — Lagrange), fueron halladas las condiciones imprescindibles y suficientes de segundo orden para dos tipos de extremo: fuerte y débil (Legendre, Jacobi y Weierstrass), así como un nuevo método de abordar los problemas de variaciones (teoría de Hamiltón — Jacobi) que permitió construir la teoría del campo (Kneser e Hilbert).

Hacia mediados de los años treinta de nuestro siglo muchos consideraban que el asunto de la teoría de los problemas extremales estaba prácticamente agotado. Pero no fue así. En 1939 vinieron a consultarse con L. V. Kantoróvich, profesor y jefe de la sección del Instituto de Matemáticas y Mecánica adjunto a la Universidad de Leningrado, los representantes de un trust, pidiéndole que les ayudara a resolver varios problemas relacionados con la producción de madera contrachapada. Después de la formalización matemática se puso de manifiesto el hecho de que la solución de dichos problemas se reduce a la búsqueda del extremo de las funciones lineales en los poliedros. Pero fue imposible examinar todos los vértices de estos últimos a causa de su enorme cantidad. Pero L. V. Kantoróvich halló otros métodos de solución y estudio de tales problemas y asimismo estableció los fundamentos de una nueva rama en la teoría de programación lineal.

Los métodos de programación lineal encontraron amplia aplicación práctica, principalmente en la economía. Por su elaboración e introducción en la economía L. V. Kantoróvich fue galardonado en 1965 con el Premio Lenin, y en 1975 (junto con el economista norteamericano T. Ch. Kupmans), con el Premio Nobel.

En los años cuarenta la teoría de problemas extremales comenzó a experimentar como un segundo nacimiento. Por una parte, entonces encontró su conclusión la teoría de programación lineal. Al referirse a este período, muchos científicos occidentales destacan el papel eminente que desempeño John von Neumann. La programación lineal estimuló el desarrollo de otras ramas de la teoría de optimización y antes que nada el análisis convexo y la programación matemática.

El análisis convexo es una parte especial de las matemáticas, donde se estudian los objetos convexos: conjuntos, funciones y problemas extremales. Sus principios fueron fundamentados por T. Minkowski en el empalme del siglo pasado y el siglo presente, pero el período de mayor desarrollo del análisis convexo corresponde a los años cincuenta y sesenta.

A finales de los años cuarenta y a principios de los años cincuenta se puso de manifiesto el hecho de que muchos problemas de control óptimo que surgían en la técnica se mantenían fuera de los marcos de las teorías elaboradas para aquel entonces. En aquellos años, en el Instituto de Matemáticas V. A. Steklov de la Academia de Ciencias de la URSS fue organizado, bajo la dirección de L. S. Pontriaguin, un seminario dedicado al análisis de tales problemas. En ese seminario intervenía, en particular, con informes acerca de ciertos problemas de la regulación automática, el destacado especialista en esa materia A. A. Feldbaum. El hablaba, entre otras cosas, acerca del problema de la parada más rápida posible de los ascensores en los pozos maestros de las minas. Este era un problema que posteriormente entró en numerosas publicaciones dedicadas al control óptimo bajo el nombre de problema elemental de la acción rápida.

L. S. Pontriaguin y sus discípulos lograron formalizar toda una serie de cuestiones relacionadas con la mayoría de los problemas técnicos actuales, y luego fue construida la teoría de esa clase de problemas, la cual recibió el nombre de teoría de control óptimo y cuyo principal resultado es el llamado principio del máximo de Pontriaguin. Por la elaboración de dicha teoría L. S. Pontriaguin y sus colaboradores V. G. Boltianski, R. V. Gamkrelidze y E. F. Mishenko fueron galardonados con el Premio Lenin en el año 1962.

Es interesante notar que el primer problema realmente relacionado con el control óptimo fue planteado por Newton en sus «*Philosophiae naturalis principia mathematica*» (año 1687) incluso antes de que apareciera la braquistócrona.

Al referirnos a las principales etapas de desarrollo de la teoría de problemas de máximo y mínimo hemos citado los problemas concretos más importantes que surgen tanto en las propias matemáticas como en las ciencias naturales, la economía y la técnica, y los cuales estimularon el desarrollo de la referida teoría *. El lector podrá enterarse más detalladamente de todo eso en las páginas de este manual.

La primera parte del libro contempla en realidad la metodología única del análisis de problemas extremales que asciende a Lagrange. Es una metodología muy simple, pero su estricta argumentación no siempre es elemental. Aprender a resolver problemas es mucho más fácil que dominar los métodos de demostración. Por eso el libro contiene el § 12 y el apartado de problemas que permiten dominar mejor dicha metodología. Además, en la primera parte del manual se ofrecen numerosos ejemplos y problemas ilustrativos que se analizan con detalles. Particularmente, en estos apartados del libro se brinda el examen minucioso de todos los problemas concretos (isoperimétricos clásicos, de braquistócrona, de Newton, de acción rápida, etc.), mencionados más arriba.

Anteriormente hemos dicho que la necesidad de resolver problemas extremales fue una de las causas que estimularon el origen y el desarrollo del análisis clásico, es decir, el cálculo diferencial e integral de funciones de un número finito de variables. Pero ya el problema de la braquistócrona no pertenece al análisis de dimensión finita. Aquí deben compararse una con otra las curvas (bastante planas) que pasan por puntos dados, o sea, un objeto de dimensión infinita.

Los primeros trabajos dedicados al cálculo de variaciones se refieren indudablemente al análisis de dimensión infinita, pero esa tendencia en las matemáticas surgió de por sí tan sólo a finales del siglo pasado y a principios de nuestro siglo.

El análisis de dimensión infinita comprende, en particular, el cálculo diferencial en espacios de dimensiones infinitas, y el mismo

---

* Al final del libro se da una breve tabla cronológica.

no es, en principio, más complejo que nuestro cálculo diferencial ordinario de dimensión finita. Además, la ecuación de Euler en el cálculo de variaciones no es nada más que el desciframiento del teorema de Fermat acerca de que la derivada es igual a cero. Y, en general, para comprender bien la unicidad de todas las partes de la teoría de problemas extremales y, en particular, de las partes que fueron descritas más arriba (programación matemática, cálculo de variaciones, programación lineal y control óptimo), lo mejor de todo es subir al nivel del análisis de dimensión infinita, el cual se alcanza hace tiempo por instrucción universitaria y en la enseñanza técnico-matemática con programas intensos de esta última asignatura.

La segunda parte del libro tiene el propósito de inculcarle al lector la idea de que la teoría de optimización es una especie de factor único. Esa unicidad se consigue agrupando el análisis plano de dimensión infinita y el análisis convexo. (A propósito, estimamos que llegó la hora en que los conceptos y hechos fundamentales del análisis convexo y de dimensión infinita han de ocupar su lugar en la enseñanza matemática general de cualquier nivel. Para contribuir a ello hemos prestado gran atención a dichas cuestiones tanto en la primera como en la segunda parte del libro).

En el § 7 se ofrecen datos acerca del análisis funcional, necesarios para construir el fragmento teórico contenido en los demás párrafos. Ese fragmento es bastante amplio y, al realizar un curso concreto, será necesario elegir alguna de sus partes. Para que el lector pueda alcanzar el objetivo necesario utilizando el camino más corto, al comienzo de la segunda parte se expone el árbol de relaciones entre los principales teoremas de los problemas extremales, juntamente con los hechos básicos del análisis y la geometría (§ 7).

Consideramos que el perfeccionamiento de la enseñanza y su modernización, en particular, la modernización de un curso independiente, ha de atravesar cierta etapa intermedia cuando se halla presente el material tradicional aprobado por muchas décadas de enseñanza y cuando existe la posible aparición de nuevos puntos de vista en cuanto a todo el tema en general. Quisiéramos que el presente manual contribuyera al perfeccionamiento de esta importante rama de la enseñanza matemática.

2. Conceptos fundamentales relacionados con los problemas extremales. La palabra *máximo* o *máximum* en latín significa «lo mas grande», y la palabra *mínimo* o *minimum*, «lo más pequeño». Estos dos conceptos — máximo y mínimo — en conjunto constituyen un término único denominado *extremo* (de la palabra latina *extremum*). A veces se utiliza la palabra *óptimo* (del latín *optimus*, que significa «lo mejor o lo más perfecto»). La teoría de problemas destinados a la búsqueda de magnitudes máximas y mínimas se denomina *teoría de problemas extremales* o *teoría de optimización*, aunque a veces también se llama *teoría de control óptimo*. Siempre que se use el último término suele suponerse que los problemas tienen aplicación práctica.

Los problemas extremales que surgen en las ciencias naturales o en la práctica se plantean, por lo común, oralmente, utilizando los términos sustanciales de la esfera económica donde surgió un problema dado. Para poder utilizar dicha teoría es necesario traducir los problemas al lenguaje matemático. Tal traducción se denomina *formalización*. Un mismo problema puede ser formalizado usando distintos métodos, y la simplicidad de su solución a menudo depende considerablemente del grado de acierto con que el mismo ha sido formalizado.

Todo problema formalizado comprende los siguientes elementos: *la funcional* $f: X \to \overline{\mathbf{R}}$ (X es el *campo de definición de la funcional* f) y la *acotación*, es decir, el subconjunto $C \subset X$.

Aclaremos algunas denotaciones y términos aquí citados: $\overline{\mathbf{R}}$ es una recta real ensanchada (sustancial), o sea, el conjunto de todos los números reales completados por los valores de $+\infty$ y $-\infty$; la notación $F: X \to Y$ significa que la aplicación F tiene el campo de definición X, mientras que $F(x)$ para cada elemento x de X se halla en el conjunto Y. Por consiguiente, la formalización de un problema extremal se reduce a la descripción exacta de sus elementos f, X y C.

Para un problema formalizado se utiliza la siguiente notación:

$$f(x) \to \inf (\sup); \quad x \in C. \tag{Pr}$$

Los puntos $x \in C$ se llaman *admisibles*. Cuando $C = X$ se dice que tenemos un *problema no acotado*.

El problema de máximo siempre se puede reducir a un problema de mínimo, sustituyendo el problema ($f(x) \to \sup$; $x \in C$ por el problema $\tilde{f}(x) \to \inf$; $x \in C$, donde $\tilde{f}(x) = -f(x)$. Y al contrario, el problema de mínimo puede ser análogamente reducido a un problema de máximo. Para asegurar la certeza de los cálculos cuando difieren las definiciones de las condiciones de extremo necesarias en los problemas de máximo y mínimo, las mismas se escriben tan sólo para los problemas de mínimo. Pero si es necesario examinar ambos problemas, entonces escribimos $f(x) \to \text{extr}$; $x \in C$.

El punto admisible $\hat{x}$ del problema (Pr) se llama *mínimo (máximo) absoluto (global)* cuando $f(x) \geqslant f(\hat{x})$ para cualquier $x \in C$ (respectivamente, $f(x) \leqslant f(\hat{x})$ para cualquier $x \in C$). En este caso escribimos $\hat{x} \in \text{absmín Pr}$ (absmáx Pr). El mínimo (máximo) absoluto se denomina *solución del problema*. La magnitud $f(\hat{x})$, donde $\hat{x}$ es la solución de este último, se llama *valor numérico del problema* (a veces, para abreviar decimos simplemente valor del problema). Esa magnitud se designa por S_{Pr} o $S_{\text{mín}}$ ($S_{\text{máx}}$).

Además de los extremos globales, también examinaremos los extremos locales dándoles una definición estricta. Sea X un espacio normalizado en el problema (Pr). Se dice que el punto $\hat{x}$ proporciona a Pr un *mínimo (máximo) local*, y se escribe $\hat{x} \in \text{locmín Pr}$ (locmáx Pr) si $\hat{x} \in C$ y si existe $\delta > 0$ tal que para cualquier punto admisible x, cuando $\|x - \hat{x}\| < \delta$, se cumple la desigualdad $f(x) \geqslant f(\hat{x})$ ($f(x) \leqslant f(\hat{x})$). Con otras palabras, si $\hat{x} \in$ locmín Pr (locmáx Pr), existirá un entorno U del punto $\hat{x}$ tal que $\hat{x} \in$ absmín Pr' (absmáx Pr') en el problema

$$f(x) \to \inf (\sup); \quad x \in C \cup U. \tag{Pr'}$$

Reformemos el concepto de extremo local en el caso más importante de dimensión finita. Sea $X = \mathbf{R}^n$ en (Pr). Dícese que $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n) \in$ locmín Pr (locmáx Pr) si $\hat{x} \in C$ y si existe $\delta > 0$ tal que para cualquier $x = (x_1, \ldots, x_n) \in C$, cuando $|x - \hat{x}| < \delta \Leftrightarrow \Leftrightarrow \left(\sum_{i=1}^{n} (x_i - \hat{x}_i)^2\right)^{1/2} < \delta$, se cumple la desigualdad $f(x) \geqslant \geqslant f(\hat{x})$ ($f(x) \leqslant f(\hat{x})$).

La teoría de problemas extremales nos brinda las reglas que permiten resolver estos problemas. En su mayoría dichas reglas forman cierto subconjunto de puntos entre los cuales debe existir la solución del problema. El referido conjunto de puntos, denominado crítico, probablemente es algo más amplio que el conjunto de extremos absolutos e incluso locales. Una vez determinados todos los puntos críticos, es necesario formar de ellos la solución.

3. Principio de Lagrange para el análisis de los problemas acotados. El principio de Lagrange presupone la reducción de problemas acotados a estructuras más simples (a problemas no acotados, en la mayoría de los casos).

Al recurrir a problemas de dimensión finita con acotaciones de tipo de igualdades, mostremos en qué consiste el principio de Lagrange.

Analicemos el problema ($X = \mathbf{R}^n$)

$$f_0(x) \to \text{extr}; \; f_1(x) = 0, \ldots, f_m(x) = 0, \tag{Pr}$$

donde $x = (x_1, \ldots, x_n)$. Aquí la acotación está representada por el sistema de igualdades $C = \{x \in \mathbf{R}^n \mid f_i(x) = 0, \, i = 1, \ldots, m\}$. La funcional f_0 y las funciones f_i que forman las ecuaciones de relación $f_i(x) = 0$ suponemos que son continuamente diferenciables (es decir, tales que todas sus derivadas parciales de primer orden son continuas).

Veamos cómo Lagrange personalmente recomendaba resolver ese problema. El escribía: «El principio general puede ser enunciado del siguiente modo. Si se busca un máximo o mínimo de cierta función de muchas variables, a condición de que entre estas últimas exista una relación dada por una o varias funciones, es necesario añadir a la función cuyo extremo se busca, las funciones que forman las ecuaciones de relación, multiplicadas por factores indefinidos, y buscar luego el máximo o mínimo de la suma construida, como si las variables fueran independientes. Las ecuaciones obtenidas, unidas a las ecuaciones de relación, servirán para definir todas las incógnitas».

Utilicemos la regla lagrangiana precisándola un poco. Lo primero que tenemos que hacer, según Lagrange: «agregar a la función cuyo extremo se busca, las funciones que forman las ecuaciones de relación, multiplicadas por factores indefinidos». Escribamos la función

$$\mathcal{L} = \mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x), \ \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$$

llamada *función de Lagrange*. Los números λ_i se llaman *factores de Lagrange*. La primera precisión consiste en que la función cuyo extremo se busca, tambien ha sido multiplicada complementariamente por un factor indefinido. Si no se hace tal precisión, la receta de Lagrange puede resultar errónea (véase a continuación el ejemplo 1). Además será necesario tomar $\lambda_0 \geqslant 0$ en el problema de mínimo y $\lambda_0 \leqslant 0$ en el de máximo.

Lo segundo que se necesita hacer siguiendo a Lagrange: «buscar el máximo o mínimo de la suma construida tomando por independientes las variables». Eso quiere decir que, según la idea de Lagrange, se ha de analizar el problema

$$\mathcal{L}(x, \lambda) \to \text{extr (por } x) \qquad (\mathrm{Pr}_{\mathrm{extr}})$$

(fijando mentalmente λ).

El problema ($\mathrm{Pr}_{\mathrm{extr}}$), perteneciendo a la clase de elementales, es más sencillo que el inicial porque carece de acotaciones. No vamos a buscar sus máximos y mínimos, porque podrá resultar que los últimos no tienen relación alguna con los máximos y mínimos del problema inicial (ejemplo 2). Procedamos de manera un poco distinta. Busquemos los *puntos estacionarios en el problema* ($\mathrm{Pr}_{\mathrm{extr}}$), o sea, anotemos, para el problema elemental, la condición necesaria de mínimo o máximo expresada en el propio teorema de Fermat. Según ese teorema, tienen que satisfacerse las ecuaciones

$$\mathcal{L}_x(x, \lambda) = 0 \iff \mathcal{L}_{x_i}(x_1, \ldots, x_n, \lambda_0, \ldots, \lambda_m) = 0, \quad i = 1, \ldots, n$$

(donde no todos los factores de Lagrange son iguales a cero). Las n ecuaciones obtenidas, unidas a η ecuaciones de relación, serán las que «servirán para la definición de todas las incógnitas». Efectivamente, si bien las incógnitas (x, λ) cuentan con una unidad más que las ecuaciones, hay que tener presente la circunstancia de que los factores lagrangianos pueden ser multiplicados por cualquier número distinto de cero. Y precisamente por eso el número de ecuaciones es igual a la cantidad de incógnitas. En estos casos dícese sobre la *totalidad del surtido de condiciones* para la definición de los puntos estacionarios. Hay que tener en cuenta que son de máximo interés los casos cuando $\lambda_0 \neq 0$, porque con $\lambda \neq 0$ las correlaciones del principio de Lagrange testimonian no más que sobre cierta degeneración de las acotaciones (fáciles de evitar, en la mayoría de los casos) y no están relacionadas con la funcional. Son precisamente las soluciones de las ecuaciones obtenidas $\mathscr{L}_{x_i} = 0$, $i = 1, \ldots, n$, $f_j(x) = 0$, $j = 1, \ldots, m$) las que forman el conjunto de puntos estacionarios.

Así pues, para resolver el problema (Pr) es necesario

1. Deducir la función de Lagrange:

$$\mathscr{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x).$$

2. Escribir las condiciones necesarias de extremo:

$$\mathscr{L}_x(x, \lambda) = 0 \Leftrightarrow \sum_{i=0}^{m} \lambda_i \frac{\partial f_i(x)}{\partial x_j} = 0, \quad j = 1, \ldots, n.$$

3. Hallar los puntos que, siendo estacionarios o admisibles, dan soluciones a las ecuaciones (p. 2), en las cuales no todos los λ_i, $i = 0, 1, \ldots, m$ son iguales a cero. En esas condiciones es útil analizar aparte los casos de $\lambda_0 = 0$ y $\lambda_0 \neq 0$. En el segundo caso, al resolver el problema de mínimo se puede tomar λ_0 por unidad o igual a cualquier otra constante positiva, y en el problema de máximo, igual a menos uno o a cualquier otra constante negativa.

4. Buscar la solución entre todos los puntos estacionarios o demostrar que no la hay.

El procedimiento descrito, llamado *principio de Lagrange*, es aplicable no sólo al problema (Pr), sino también a un amplio grupo de problemas extremales. La mayoría de los problemas que se citan en nuestro libro pueden resolverse con ayuda del principio en cuestión. Pero aquí es importante tener en cuenta lo siguiente.

a) Hablando en general, el principio de Lagrange no siempre es aplicable. El problema que citamos a continuación (ejemplo 3) tiene solución, pero el principio de Lagrange no conduce a ella.

b) La esfera de aplicabilidad del principio de Lagrange es bastante extensa. No se aplica, a veces, al problema algún teorema disponible, mientras que dicho principio (aplicado sin argumentaciones) lleva, no obstante, a puntos propensos al extremo, dando la posibilidad de obtener de ellos la solución.

En un caso general el principio de Lagrange se aplica de la siguiente manera.

1. Formalizar el problema a la especie de

$$f(x, u) \to \inf; \quad F(x, u) = 0, \quad u \in U.$$

$f : X \times U \to \overline{\mathbf{R}}$, $F : X \times U \to Y$, donde X e Y son espacios normalizados. Aquí se trata del problema con acotaciones del tipo de igualdades que son parametrizadas por cierto conjunto U. Se puede decir también que es un *problema con acotaciones del tipo de igualdades e inclusiones.*

Deducir la función de Lagrange:

$$\mathcal{L}(x, u, \lambda_0, y^*) = \lambda_0 f(x, u) + \langle y^*, F(x, u)\rangle,$$

donde y^* es el elemento del espacio conjugado Y^*.

La función de Lagrange carece de acotaciones del tipo de inclusiones $u \in U$.

2. Anotar las condiciones necesarias para los problemas

$$\mathcal{L}(x, \hat{u}, \lambda_0 y^*) \to \inf \text{ (por } x)$$

$$\mathcal{L}(x, \hat{u}, \lambda_0, y^*) \to \inf, \ u \in U.$$

3. Hallar los puntos *críticos*, o sea, admisibles, que, siendo solución de las ecuaciones del p. 2, donde y^* y λ_0 no son iguales simultáneamente a cero. Sería más conveniente aquí analizar aparte los casos de $\lambda_0 = 0$ y $\lambda_0 \neq 0$. En el segundo caso se puede tomar λ_0 por unidad o igual a cualquier otra constante positiva.

4. Buscar soluciones entre todos los puntos críticos o demostrar que no las hay.

En resumen citemos los tres ejemplos sobre los cuales hemos hablado más arriba.

Ejemplo 1 (evidencia que no siempre se estima $\lambda_0 = 1$ en la regla de factores de Lagrange).

$$f_0(x_1, x_2,) = x_1 \to \inf; \quad f_1(x_1, x_2) = x_1^3 - x_2^2 = 0 \qquad (X = \mathbf{R}^2).$$

Las funciones f_0 y f_1 son continuamente diferenciables. Es fácil comprender que la solución del problema es $\hat{x} = (0, 0)$. Si seguimos directamente a Lagrange, debemos escribir la suma $\mathcal{L} = x_1 + \lambda(x_1^3 - x_2^2)$ y resolver luego las ecuaciones

$$\mathcal{L}_{x_1} = 0, \quad \mathcal{L}_{x_2} = 0 \Leftrightarrow 1 + 3\lambda x_1^2 = 0, \quad -2\lambda x_2 = 0.$$

Pero esas ecuaciones son incompatibles con la ecuación de relación $x_1^2 - x_2^2 = 0$.

Ejemplo 2 (evidencia que el extremo de la función de Lagrange, como problema no acotado, puede no coincidir con el extremo de la versión inicial acotada).

$$f_0(x_1, x_2) = x_2^2 - x_1 \to \inf;$$
$$f_1(x_1, x_2) = x_1 x_1^3 = 0 \qquad (X = \mathbf{R}^2).$$

Claro que la solución del problema es $\hat{x}$ (0, 0). La función de Lagrange: $\mathcal{L} = \lambda_0 (x_2^2 - x_1) + \lambda(x_1 + x_1^3)$. La condición necesaria de extremo:

$$\mathcal{L}_{x_1} = 0, \quad \mathcal{L}_{x_2} = 0 \Leftrightarrow -\lambda_0 + \lambda(1 + 3x_1^2) = 0, \quad 2\lambda_0 x_2 = 0.$$

Si $\lambda_0 = 0$, entonces $\lambda \neq 0$ y, consiguientemente, $1 + 3x_1^2 = 0$ será la contradicción. Por lo tanto, $\lambda_0 \neq 0$. Supongamos que $\lambda_0 = 1$. Entonces la función de Lagrange adoptará el siguiente aspecto:

$$\mathcal{L} = x_2^2 - x_1 + \lambda(x_1 + x_1^3).$$

Sin embargo, cualesquiera que sean λ, esa función no tendrá en el punto $\hat{x} = (0, 0)$ incluso mínimo local.

Ejemplo 3 (evidencia que si no se observan determinadas condiciones, el principio de Lagrange conduce a resultados erróneos). Sea

$$X = Y = l_2,\ f(x) = x_1 + x_2/2 + \ldots + x_n/n + \ldots,$$
$$F(x) = (x_1,\ x_2/2,\ \ldots,\ x_n/n,\ \ldots)$$
$$(x = (x_1, x_2, \ldots, x_n, \ldots)).$$

Analicemos el problema

$$f(x) \to \inf;\ F(x) = 0.$$

Aquí $\hat{x} = 0$ es el único elemento admisible y representa, por consiguiente, precisamente la solución del problema. Si suponemos que existen los factores de Lagrange $\lambda_0 \in \mathbf{R}$ e $y^* \in l_2$ no iguales simultáneamente a cero y tales que, para el problema elemental $\mathcal{L} = \lambda_0 f(x) + \langle y^*, F(x)\rangle \to \inf$, sea cumplida la condición necesaria de mínimo $\mathcal{L}(x, \lambda_0\, y^*)$ (teorema de Fermat), entonces

$$\mathcal{L}_x(x, \lambda_0 y^*) = 0 \Leftrightarrow \lambda_0 = -y_1, \ldots, \lambda_0 = -y_n, \ldots,$$

donde $y^* = (y_1, \ldots, y_n \ldots) \in l_2$, puesto que l_2^* es isomorfo a l_2 [KF, pág. 177]. Pero dichas condiciones son contradictorias: o bien $\lambda_0 \neq 0$, entonces $y^* = (-\lambda_0, \ldots, -\lambda_0, \ldots) \in l_2$; o bien $\lambda_0 = 0$, entonces $y^* = 0$, es decir, los dos factores de Lagrange son iguales a cero. Aquí l_2 es el espacio de todas las sucesiones $x = (x_1, \ldots, \ldots, x_n, \ldots)$ para las cuales $\| x \|_{l_2} = \left(\sum_{n=1}^{\infty} |x_n|^2 \right)^{1/2} < \infty$; y l_2, el espacio conjugado con l_2 [KF, pág. 177].

Ejercicios

Citar en los ejercicios 1—3 los problemas de extremo no acotados de las funciones infinitamente diferenciables o una o dos variables, en los cuales se cumplen las condiciones expuestas a continuación:

1. Hay un sólo extremo local que no es global.
2. Hay un número infinito de máximos locales, pero no hay ni un sólo mínimo local.
3. La acotación de la función dada en el plano, respecto a cualquier recta que atraviesa el origen de coordenadas, tiene en cero un mínimo local, pero a su vez, el origen de coordenadas no es un punto de mínimo local.
4. ¿Será posible afirmar que si la función de una variable tiene un mínimo local en cualquier punto, en cierto entorno bastante pequeño de dicho punto esa función decrecerá a la izquierda y crecerá a la derecha?
5. Supongamos que la función f una vez definida y diferenciable en $\mathbf{R}^n$, satisface la condición $\lim\limits_{|x| \to \infty} f(x) = +\infty$ y $f'(x)$ tiene el único cero $\hat{x}$. Demostrar que $\hat{x}$ es un punto del mínimo absoluto de la función f.
6. Supongamos que cada funcional alcanza su mínimo absoluto en cierto conjunto X. Demostrar que X es un conjunto finito.

PRIMERA PARTE

PRINCIPALES ELEMENTOS DE LA TEORÍA DE PROBLEMAS EXTREMALES

§ 1. Elementos del análisis funcional, cálculo diferencial y análisis convexo

Las construcciones de la teoría general de problemas extremales, así como de las partes fundamentales del cálculo de variaciones y del control óptimo, requieren examen en los espacios normalizados y banachianos. No obstante, los fragmentos más importantes de dicha teoría pueden ser considerados en el espacio euclidiano $\mathbf{R}^n$. En la primera parte del presente libro nos limitaremos, donde sea posible, al caso de dimensión finita. La información necesaria de todos los datos referentes al análisis funcional, al cálculo diferencial y al análisis convexo (con demostraciones completas) se ofrece en la segunda parte del libro (§§ 7 y 8). Aquí usamos no más que los elementos imprescindibles para el análisis de dimensión infinita, prestando especial atención al caso elemental de dimensión finita.

1.1. Espacios normalizados y banachianos

1.1.1. Espacio $\mathbf{R}^n$. $\mathbf{R}^n$ es el espacio de los vectores n-dimensionales $\mathbf{R}^n = \{x = (x_1, \ldots, x_n) \mid x_i \in \mathbf{R},\ i = 1, \ldots, n\}$. Los vectores de $\mathbf{R}^n$ pueden ser sumados (según la regla: $x = (x_1, \ldots, x_n)$, $x' = (x_1', \ldots, x_n') \Rightarrow x_1 + x_1' = (x_1 + x_1', \ldots, x_n + x_n')$) y multiplicados por números reales (según la regla: $x = x_1, \ldots, x_n$). $\lambda \in \mathbf{R} \Rightarrow \lambda x = (\lambda x_1, \ldots, \lambda x_n)$). El vector $0 = (0, \ldots, 0)$ se llama *vector nulo*. Eso convierte $\mathbf{R}^n$ en un *espacio lineal*.

Para introducir la distancia en el espacio $\mathbf{R}^n$ comúnmente se procede de la manera siguiente:

$$x = (x_1, \ldots, x_n), \quad x' = (x_1', \ldots, x_n') \Rightarrow d(x, x') = \sum_{i=1}^{n} (x_i - x_i')^2)^{1/2}.$$

Esa distancia, llamada *euclidiana*, puede ser expresada por la norma euclidiana:

$$|x| := (\sum_{i=1}^{n} x_i^2)^{1/2}$$

según la fórmula $d(x, x') = |x - x'|$.

La norma euclidiana satisface, lo cual se puede comprobar fácilmente, las siguientes relaciones:

a) $|x| \geqslant 0$ para todos $x \in \mathbf{R}^n$ y $|x| = 0 \Rightarrow x = 0$;

b) $|\alpha x| = |\alpha| \cdot |x|$ para cualquier x de $\mathbf{R}^n$ y cualquier α de $\mathbf{R}$;

c) $|x - x'| \leqslant |x| + |x'|$ para cualesquiera x y x' de $\mathbf{R}^n$.

Las funciones que poseen esas propiedades se llaman *normas*, y los espacios lineales dotados de normas se llaman *normalizados* (las definiciones exactas se dan en el punto siguiente). Las normas de $\mathbf{R}^n$ pueden ser definidas de modo diferente. Por ejemplo, a base de la fórmula

$$\| x \|_p = \left(\sum_{i=1}^{n} | x_i |^p \right)^{1/p} \quad (1 \leqslant p \infty)$$

o mediante la fórmula

$$\| x \|_\infty = \max_{1 \leqslant i \leqslant n} | x_i |.$$

La norma permite hallar la distancia por medio de la fórmula $d(x, x') = \| x - x' \|$ y definir los conjuntos abiertos, cerrados, convergentes, etc.

1.1.2. Definición y ejemplos de espacios banachianos. El espacio lineal X se llama *normalizado* si en X ha sido determinada la funcional $\| \cdot \| : X \to \mathbf{R}$ que, llamada *norma*, satisface las condiciones siguientes:

a) $\| x \| \geqslant 0 \ \forall x \in X$ y $\| x \| = 0 \Leftrightarrow x = 0$;

b) $\| \alpha x \| = | \alpha | \, \| x \| \ \forall \alpha \in \mathbf{R}, \ \ \forall x \in X$;

c) $\| x_1 + x_2 \| \leqslant \| x_1 \| + \| x_2 \| \ \forall x_1, \ x_2 \in X$.

A veces, para subrayar que la norma se da precisamente para X, escríbese $\| \cdot \|_X$. Dos normas en x $\| \cdot \|_1$ y $\| \cdot \|_2$ se denominan *equivalentes* si existen tales constantes positivas C_1 y C_2 que

$$C_1 \| x \|_1 \leqslant \| x \|_2 \leqslant C_2 \| x \|_1 \ \forall x \in X.$$

Todo espacio normalizado se transforma en *métrico* si se le introduce la distancia $d(x_1, x_2) = \| x_1 - x_2 \|$.

La sucesión $\{x_n\}$ de puntos del espacio métrico X se llama *fundamental* si satisface el *criterio de Cauchy*, es decir, si para cualquier $\varepsilon > 0$ existe tal número N_ε que $d(x_{n'}, x_{n''}) < \varepsilon$ para todos $n' > N_\varepsilon$, $n'' > N_\varepsilon$.

Si toda sucesión fundamental converge en el espacio X, este último se llama *pleno*.

El espacio pleno respecto a la distancia introducida se llama espacio *banachiano*.

Ejemplos de espacios banachianos.

1. El espacio de dimensión finita l_p^n integrado por los vectores $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ con la norma $\| x \|_p = \left(\sum\limits_{i=1}^{n} | x_i |^p \right)^{1/p}$, $1 \leqslant p \leqslant \infty$.

2. El espacio $C(K, \mathbf{R}^n)$ de los vectores funciones continuos $x(\cdot) : K \to \mathbf{R}^n$ dados en el compacto K con la norma $\| x(\cdot) \|_0 = \max\limits_{t \in K} x(t)$.

3. El espacio $C^r([t_0, t_1], \mathbf{R}^n)$ de los vectores funciones r veces continuamente diferenciados $x(\cdot) : [t_0, t_1] \to \mathbf{R}^n$, dados en el segmento finito $[t_0, t_1] \subset \mathbf{R}$, con la norma

$$\| x(\cdot) \|_r = \max \{\| x(\cdot) \|_0, \ldots, x^{(r)}(\cdot) \|_0\}.$$

4. El espacio l_2 integrado por las sucesiones $x = (x_1, \ldots, x_n, \ldots)$, para las que $\sum\limits_{i=1}^{\infty} x_i^2 < \infty$ con la norma dada por la fórmula $\| x \| = \left(\sum\limits_{i=1}^{\infty} x_i^2 \right)^{1/2}$.

1.1.3. Producto de los espacios. Sean X e Y ciertos espacios normalizados. El producto cartesiano de $X \times Y$ puede ser transformado en un espacio normalizado introduciéndole la norma

$$\| (x, y) \|_{X \times Y} = \max \{\| x \|_X, \| y \|_Y\}$$

(es fácil comprobar que todos los axiomas de la norma se cumplen). También son posibles otras normalizaciones.

Señalemos la afirmación evidente: el *producto cartesiano de los espacios banachianos es un producto banachiano.*

1.1.4. Espacio conjugado y operador conjugado. El conjunto X^* de todas las funcionales lineales continuas en X forma un espacio *conjugado* con X. A su vez, el mismo es un espacio banachiano con respecto a la norma $\| x^* \|_{X^*} = \sup\limits_{\| x \|_X \leq 1} \langle x^*, x \rangle$, donde $\langle x^*, x \rangle$ significa el efecto que ejerce sobre x la funcional x^* [KF, pág. 171]. El espacio conjugado con el espacio de dimensión finita $\mathbf{R}^n$ es isomorfo a $\mathbf{R}^n$. El producto escalar de dos vectores $y = (y_1, \ldots, y_n) \in \mathbf{R}^n$ y $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ está representado por la suma $\langle y, x \rangle = \sum\limits_{i=1}^{n} y_i, x_i$. Asimismo la suma $\sum\limits_{i=1}^{n} y_i x_i$ podrá ser designada simplemente por yx si $y \in \mathbf{R}^n$, $x \in \mathbf{R}^n$. En este caso x debe tomarse por columna e y por renglón (y entonces yx será nada más que el producto de las matrices).

Sean X e Y ciertos espacios normalizados; y $\Lambda \in \mathcal{L}(X, Y)$, el operador lineal continuo de X en Y. Entonces se puede determinar el *operador conjugado* $\Lambda^* : Y^* \to X^*$, de tal manera que $\langle y^*, \Lambda x \rangle = \langle \Lambda^* y^*, x \rangle \; \forall x \in X$ [KF, pág. 217].

Para una funcional lineal continua en el producto de espacios tiene lugar el siguiente lema evidente.

Lema. *Representemos unívocamente cualquier funcional* $\Lambda \in (X \times Y)^*$ *en forma de*

$$\langle \Lambda, (x, y) \rangle = \langle x^*, x \rangle + \langle y^*, y \rangle,$$

donde $x^* \in X^*$ e $y^* \in Y^*$.

Ejercicios

1. ¿Cuáles de las funciones de dos variables enumeradas más abajo y con qué valores paramétricos proporcionan la norma en $\mathbf{R}^2$:
a) $N(x) = (|x_1|^p - |x_2|^p)^{1/p}$, $p > 0$;
b) $N(x) = |a_{11}x_1 + a_{12}x_2| + |a_{21}x_1 + a_{22}x_2|$;
c) $N(x) = \text{máx}\{|a_{11}x_1 + a_{12}x_2|, |a_{21}x_1 + a_{22}x_2|\}$;
d) $N(x) = (a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2)^{1/2}$?
2. Demostrar que las normas $|x| = (x_1^2 + x_2^2)^{1/2}$ y $\|x\|_\infty = \text{máx}\{|x_1|, |x_2|\}$ son equivalentes.
3. Demostrar que si $\|\cdot\|$ es una norma en $\mathbf{R}^n$, la esfera unidad $B = \{x \in \mathbf{R}^n \mid \|x\| \leqslant 1\}$ en esa norma será un conjunto central simétrico, cerrado, convexo y acotado, cuyo centro (origen de coordenadas) es un punto interno.
4. Demostrar que si B es un conjunto central simétrico, cerrado, convexo y acotado en $\mathbf{R}^n$, cuyo centro (origen de coordenadas) es un punto interno, existirá una norma tal con la que B será una esfera unidad.
5. Demostrar que todas las normas en $\mathbf{R}^n$ son equivalentes.
6. Demostrar que todos los espacios normalizados de dimensiones finitas son banachianos.
7. Construir un ejemplo de espacio normalizado pero no banachiano.
8. Demostrar que si $(X, \|\cdot\|_X)$ e $(Y, \|\cdot\|)$ son espacios normalizados, en este caso $\|x\|_X + \|y\|_Y$ y $(\|x\|_X^2 + \|y\|_Y^2)^{1/2}$ serán las normas equivalentes en $X \times Y$.
9. Supongamos que el espacio normalizado X conste de funciones continuas en el segmento $[0, 1]$ con la norma $\|x(\cdot)\| = \int_0^1 x(t)\,dt$. ¿Pertenecerá la funcional lineal $\langle x^*, x(\cdot)\rangle = x(0)$ al espacio X^*?
10. ¿A qué será igual la norma $\|x\|$, $x \in \mathbf{R}^2$ si la esfera unidad ha sido deducida por las desigualdades
a) $B = \{(x_1, x_2) \mid -a_1 \leqslant x_1 \leqslant a_1, \ -a_2 \leqslant x_2 \leqslant a_2\}$;
b) $B = \left\{(x_1, x_2) \,\middle|\, \frac{x_1^2}{a_1^2} + \frac{x_2^2}{x_2^2} \leqslant 1, \ -b_1 \leqslant x_1 \leqslant b_1; \ -b_2 \leqslant x_2 \leqslant b_2\right\}$?
11. Citar un ejemplo de subespacio bidimensional $C([0, 1])$ cuya esfera unidad es un círculo unidad (con otras palabras: cortar con un plano la esfera unidad del espacio $C([0, 1])$, de tal modo que en la sección se forme un círculo).
12. Sea $X = \mathbf{R}^n$. Hallar la norma del espacio conjugado con $(X, \|x\|)$ si la norma en X se deduce de las relaciones
a) $N(x) = (|x_1|^2 + |x_2|^2)^{1/2}$;
b) $N(x) = \text{máx}(|x_1|, |x_2|)$;
c) $N(x) = (|x_1|^p - |x_2|^p)^{1/p}$, $p > 1$;
d) $N(x) = (a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2)^{1/2}$, $a_{11} > 0$, $a_{11}a_{22} - a_{12}^2 > 0$.

1.2. Algunos teoremas de geometría y análisis funcional

1.2.1. Teoremas de Weierstrass sobre el alcance del máximo y el mínimo. En la mayoría de los casos utilizaremos el siguiente principal

Teorema de Weierstrass. *La función continua que se halla en un subconjunto no vacío cerrado acotado de un espacio de dimensión finita alcanza sus máximo y mínimo absolutos* [15, t. 1, pág. 235].

Deduzcamos de este teorema un simple corolario que utilizaremos muy a menudo.

Corolario. *Si la función f es continua en $\mathbf{R}^n$ y $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$ ($\lim\limits_{\|x\| \to \infty} f(x) = -\infty$), en este caso f alcanzará su mínimo (máximo) absoluto en cualquier subconjunto cerrado de $\mathbf{R}^n$.*

Recordemos que el conjunto A en un espacio métrico se llama *compacto* si de cualquier sucesión de elementos de A se puede elegir una sucesión contenida convergente a un elemento de A, o bien si (que es lo mismo) de cualquier recubrimiento de A por conjuntos abiertos se puede elegir un subrecubrimiento finito. El subconjunto acotado y cerrado del espacio de dimensión finita es un compacto.

1.2.2. Teoremas de separabilidad. Al deducir las condiciones necesarias de extremo (principio de Lagrange) en los problemas convexos (p. 2.3.3) y en los problemas con igualdades y desigualdades (p. 2.4), aprovechamos la propiedad de separabilidad de los conjuntos convexos no intersecables. Citemos los resultados referentes al caso de dimensión finita (en el p. 7.1 se examina el caso de dimensión infinita).

Los conjuntos A y B se denominan *separables* si existe un semiespacio que contenga A, y que los puntos del conjunto B no le pertenezcan o se hallen situados en la frontera de dicho semiespacio.

Analíticamente eso se puede escribir así: existe un vector $\lambda \in \mathbf{R}^n$ $(\lambda \neq 0)$ para el cual

$$\inf_{x \in A} \langle \lambda, x \rangle \geqslant \sup_{x \in B} \langle \lambda, x \rangle.$$

Los conjuntos A y B se llaman *estrictamente separables* si existe un semiespacio tal que contenga el conjunto A y no contenga los puntos de B.

Analíticamente eso se puede escribir así: existe un vector $\lambda \in \mathbf{R}^n$ para el cual

$$\inf_{x \in A} \langle \lambda, x \rangle > \sup_{x \in B} \langle \lambda, x \rangle.$$

Teorema 1 (es el primer teorema de separabilidad para el caso de dimensión finita). *Sean A y B ciertos conjuntos convexos no vacíos en $\mathbf{R}^n$, $A \cap B = \varnothing$. Entonces los conjuntos A y B serán separables.*

Teorema 2 (es el segundo teorema de separabilidad para el caso de dimensión finita). *Sea A un conjunto cerrado convexo no vacío en $\mathbf{R}^n$, $b \notin A$. Entonces el punto b se podrá separar estrictamente de A.*

$\triangle$ A) **Demostración del teorema 2.** Analicemos el problema sobre la distancia desde el punto b hasta el conjunto cerrado A:

$$|b - x| \to \inf; \quad x \in A. \tag{Pr}$$

Dado que la función $f(x) = |b - x|$ es continua y $\lim\limits_{|x| \to +\infty} f(x) = +\infty$, según el corolario del teorema de Weierstrass (p. 1.2.1), la solución del problema (Pr) existe. Designémosla por $\hat{x}$. Está claro que $\hat{x} \neq b$ (fig. 1).

Tracemos por el punto $\hat{x}$ el hiperplano $H = \{x \in \mathbf{R}^n \mid \langle \lambda, x \rangle = \hat{\alpha}\}$ perpendicularmente a la recta $b\hat{x}$. Mostremos que el conjunto A y el punto b se hallan en semiespacios opuestos. Si algún punto $a \in A$ se hallara en el semiespacio abierto que contiene el punto b, en virtud de la convexidad de A resultaría que el segmento $[a, \hat{x}] \in A$. En este caso existiría un punto de ese segmento, perteneciente al interior de la esfera $B(b, | b - \hat{x} |)$, lo cual contradice al hecho de que $\hat{x}$ es el punto de A más cercano a b.

El hiperplano estrictamente divisor está construido. Sin acotar la generalidad podemos estimar que $\langle \lambda, b \rangle < \alpha$. Entonces $\langle \lambda, a \rangle \geqslant \alpha$ para cualquier $a \in A$.

B) Demostración del teorema 1. La afirmación de que «los conjuntos A y B son separables» equivale a la afirmación de que «el punto cero es separable del conjunto $C \overset{\text{def}}{=} A - B = \{x = a - b \mid a \in A, b \in B\}$». Efectivamente,

$$\inf_{a \in A} \langle \lambda, a \rangle \quad \sup_{b \in B} \langle \lambda, b \rangle \Leftrightarrow \inf_{a \in A} \langle \lambda, a \rangle - \sup_{b \in B} \langle \lambda, b \rangle \geqslant 0 \Leftrightarrow \inf_{a \in A} \langle \lambda, a \rangle + \inf_{b \in B} \langle \lambda, -b \rangle \geqslant 0 \Leftrightarrow \inf_{\substack{a \in B \\ b \in B}} \langle \lambda, a - b \rangle \geqslant 0 = \{\lambda, 0\}.$$

Por eso para demostrar el teorema es suficiente probar que el punto cero es separable del conjunto C. Está claro que C es un conjunto convexo y que $0 \notin C$.

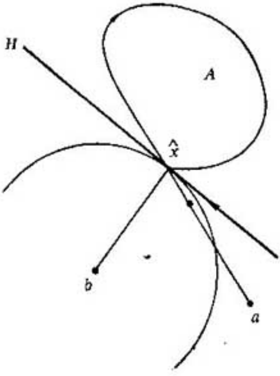


Fig. 1

Si $0 \in \overline{C}$ (clausura del conjunto C), la separabilidad se deduce del teorema 2 ya demostrado. Por eso de aquí en adelante consideraremos que $0 \in \overline{C}$. Por lo tanto, debemos demostrar la existencia de un hiperplano de apoyo para el conjunto convexo, el cual pase por el punto situado en su frontera.

Supongamos que $\inf C = \varnothing$. Entonces la dimensión de la envoltura afín del conjunto C será inferior a n, es decir, $\dim \mathit{aff}\, C < n$. En vista de que $0 \in \overline{C}$, resulta que $0 \in \mathit{aff}\, C$. En este caso $\mathit{aff}\, C$ es el propio subespacio en $\mathbf{R}^n$, y por eso existe el hiperplano $\langle \lambda, x \rangle = 0 \; \forall x \in C$ que lo contiene. Dicho hiperplano es precisamente el que se busca.

Sea $\operatorname{int} C \neq \varnothing$, es decir, existe el punto $y \in \operatorname{int} C$ $(y \neq 0)$ y, por consiguiente, cierta esfera $B_y \subset C$. Entonces, el punto $-y$ y, simétrica a éste, la esfera B_{-y} con el centro en el punto $-y$, no pertenecen a un cono $K = \{x = \sum_{i=1}^{n} t_i c_i,\ c_i \in C,\ t_i \geqslant 0,\ i = 1, \ldots, s,\ s \in N$ cualquiera$\}$, que es (como se puede comprobar fácilmente) un conjunto convexo. De lo contrario resultaría que $0 \in C$. Así pues, $y \notin \overline{K}$ corresponde a un conjunto convexo cerrado.

Con arreglo al segundo teorema de separabilidad, el punto $-y$ y el conjunto $\overline{K}$ pueden ser estrictamente separados, es decir, existe un vector $\lambda \neq 0$ para el cual

$$\inf_{x \in K} \langle \lambda, x \rangle > \langle \lambda, -y \rangle. \tag{1}$$

Es imposible admitir que para cierto punto $x \in \overline{K}$ se cumpla la desigualdad $\langle \lambda, x \rangle < 0$. De ser así, la cota inferior del primer miembro de la desigualdad (1) sería igual a $-\infty$. Por consiguiente, $\langle \lambda, x \rangle \geqslant 0 \forall x \in \overline{K}$, además, $\forall x \in C$. Así pues, el hiperplano que divide el punto cero y el conjunto C ya está construido.

Ejercicios

1. Citar un ejemplo de función continua acotada en el subconjunto acotado de una recta para la cual no se alcanzan las cotas inferior y superior.

2. Citar un ejemplo de función continua acotada en el subconjunto cerrado de una recta para la cual no se alcanzan las cotas superior e inferior.

3. Sea X cierto subconjunto de una recta, el cual no es compacto. Demostrar que existe una función continua en X, cuya cota inferior no se puede alcanzar.

4. Averiguar si son compactos los siguientes conjuntos:

a) el semiintervalo $[a, b]$;

b) la sucesión de puntos en la recta $x_n = \frac{1}{n}$, $n = 1, 2, \ldots$;

c) el subconjunto de la recta $B = \bigcup_{n=1}^{\infty} \left[n, n + \frac{1}{n}\right]$;

d) el elipsoide $E - \{x = \{x_k\} k \geqslant 1 \in l_2 \,\Big|\, \sum_{k=1}^{\infty} k^2 x_k^2 \leqslant 1$ en el espacio l_2.

5. Citar un ejemplo de conjunto acotado cerrado que no es compacto.

6. Citar un ejemplo de espacio normalizado X y de funcional continua $f: X \to \mathbb{R}$ tal que $f(x) \to +\infty$ para $\|x\| \to \infty$ cuando su cota inferior no se puede alcanzar.

7. Citar un ejemplo de espacio banachiano X, de su subespacio cerrado L y del punto $\hat{x}$ no perteneciente a ese subespacio, tales que no tenga solución el problema de la distancia más corta desde ese punto hasta el referido subespacio.

8. Separar el punto (2,3) del elipsoide $x^2/4 + y^2/9 = 1$.

1.3. Definiciones de las derivadas

Para las funciones reales de una variable real se emplean dos definiciones:

la existencia de límite finito

$$\lim_{h \to \lambda} \frac{F(\hat{x} - h) - F(\hat{x})}{h} \tag{1}$$

y la posibilidad de desarrollo asintótico con $h \to 0$

$$F(\hat{x} + h) = F(\hat{x}) + F'(\hat{x})\, h - o(h), \tag{2}$$

que conducen a un mismo concepto de diferenciabilidad. Pero ya para las funciones de dos variables y más, existen varios criterios distintos al concepto de diferenciabilidad (suavidad). La definición (1) conduce a conceptos de derivada direccional, de variación de Lagrange y de derivada de Gateaux. La definición (2) conduce a los conceptos de derivada de Fréchet y de diferenciabilidad estricta.

1.3.1. Derivada direccional, variación de Lagrange y derivada de Gateaux. Sean X, Y ciertos espacios normalizados, y $F : X \to Y$, la aplicación del espacio X o de cierto entorno del punto $\hat{x}$ en el espacio Y. Suponiendo que existe el límite

$$\delta F(\hat{x}, h) = \lim_{\lambda \to 0} \frac{F(\hat{x} + \lambda h) - F(\hat{x})}{\lambda},$$

el mismo se llama *derivada F en el punto $\hat{x}$ de la dirección h*.

Si la aplicación F dispone en el punto $\hat{x}$ de una derivada en todas las direcciones $h \in X$, dícese que F tiene en el *punto $\hat{x}$ variación de Lagrange*. En este caso la aplicación $h \to \delta F(\hat{x}, h)$ se denomina *variación de Lagrange*.

Si el operador $\delta F(\hat{x}, \cdot) : X \to Y$ es lineal y continuo en h, dícese que *F es diferenciable según Gateaux en el punto $\hat{x}$*, mientras que el operador $\delta F(\hat{x}, \cdot)$, llamado *derivada de Gateaux de la aplicación F en el punto $\hat{x}$*, se designa por $F'_G(\hat{x})$. Por lo tanto, siempre que F sea diferenciable según Gateaux en el punto $\hat{x}$, para cualquier h fija tendrá lugar el desarrollo

$$F(\hat{x} + \lambda h) = F(\hat{x}) + \lambda F'_G(\hat{x})[h] + r(h, \lambda),$$

donde $\| r(h, \lambda) \| \to 0$ para $\lambda \to 0$.

1.3.2. Derivada de Fréchet y diferenciabilidad estricta. La aplicación F (p. 1.3.1) se llama *diferenciable según Fréchet en el punto $\hat{x}$* y se escribe $F \in D(\hat{x})$ si existe un operador lineal continuo de X en Y, designado por $F'(\hat{x})$, y una aplicación r de cierto entorno de $\hat{x}$ en Y, tales que

$$F(\hat{x} + h) = F(\hat{x}) + F'(\hat{x})[h] + r(h), \qquad (1)$$
$$\| r(h) \| = 0(\| h \|) \text{ para } \| h \| \to 0.$$

El operador $F'(\hat{x}) : X \to Y$ se llama *derivada de Fréchet*. Brevemente la correlación (1) se puede escribir así:

$$F(\hat{x} + h) = F(x) + F'(\hat{x})[h] + o(h),$$

considerando $o(h)$ como un elemento del espacio Y, para el cual $\| o(h) \| = o(\| h \|)$ cuando $\| h \| \to 0$. Con $F'(\hat{x})[h]$ se denota el valor de la aplicación $F'(\hat{x})$ en el elemento $h \in X$. Si en cada punto x del conjunto abierto U, las aplicaciones $F \in D(x)$ y $x \to F'(x)$ son continuas, entonces se escribe $F \in C^1(U)$. Está claro que de la diferenciabilidad de Fréchet de la aplicación F en el punto $\hat{x}$ se deduce la diferenciabilidad de Gateaux de la aplicación F en ese mismo punto. Pero ya en el caso bidimensional esos dos conceptos se diferencian. De la diferenciabilidad de Gateaux, según la definición, se desprende la existencia de la variación de Lagrange. Y otra vez (para el caso bidimensional) esos conceptos son diferentes.

La definición de la diferenciabilidad de Fréchet de la aplicación F en el punto $\hat{x}$, en el lenguaje ε-δ se enuncia así: existe un operador $F'(\hat{x}) \in \mathcal{L}(X, Y)$ tal que para cualquier $\varepsilon > 0$ se encontrará $\delta > 0$ con el que se cumpla, para todas h: $\| h \| < \delta$, la desigualdad

$$\| F(\hat{x} + h) - F(\hat{x}) - F'(x)[h] \| \leqslant \varepsilon \| h \|.$$

De la expresión (1) se deduce que la derivada de Fréchet está unívocamente definida, ya que la igualdad $\Lambda_1 h - \Lambda_2 h = o(h)$ para los operadores lineales continuos Λ_1 y Λ_2 será posible tan sólo cuando $\Lambda_1 = \Lambda_2$.

En muchos problemas el análisis de dimensión finita y de dimensión infinita, de diferenciabilidad de Fréchet en un punto, no es suficiente para obtener un resultado satisfactorio. Esa circunstancia obliga a intensificar aún más la diferenciabilidad en el punto.

Supongamos que la aplicación F es diferenciable según Fréchet en el punto $\hat{x}$. Esa aplicación se llama *estrictamente diferenciable en el punto* $\hat{x}$ (además se escribe $F \in SD(\hat{x})$) si para cualquier $\varepsilon > 0$ se halla $\delta > 0$ tal que para todos los valores de x_1 y x_2 que satisfagan las desigualdades $\| x_1 - \hat{x} \| < \delta$, $\| x_2 - \hat{x} \| < \delta$ se cumpla la desigualdad

$$\| F(x_1) - F(x_2) - F'(\hat{x})[x_1 - x_2] \| \leqslant \varepsilon \| x_1 - x_2 \|.$$

Si $F: \mathbf{R}^n \to \mathbf{R}^m$ es la aplicación de un espacio de dimensión finita $\mathbf{R}^n$ diferenciable en el punto $\hat{x}$ o de cierto entorno del punto $\hat{x}$ en el espacio $\mathbf{R}^m$, la derivada de Fréchet en $\hat{x}$ será una matriz integrada por las derivadas parciales

$$F'(\hat{x}) = \left(\frac{\partial F_i(\hat{x})}{\partial x_j}\right), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

La misma se llama *matriz de Jacobi*. Si $m = n$, el determinante de dicha matriz se denomina *determinante jacobiano* de la aplicación F en el punto $\hat{x}$.

1.3.3. Derivadas parciales y derivadas de órdenes superiores. Sean X, Y, Z ciertos espacios normalizados, y U, cierto entorno del punto $(\hat{x}, \hat{y})$ en $X \times Y$, $F: U \to Z$.

Si la aplicación $x \to F(x, \hat{y})$ es diferenciable según Fréchet en el punto $\hat{x}$, su derivada designada por $F'_x(\hat{x}, \hat{y})$ o $\frac{\partial F(\hat{x}, \hat{y})}{\partial x}$ se llamará *devarida parcial en* x de la aplicación F al punto $(\hat{x}, \hat{y})$.

Análogamente se define la derivada parcial en y:

$$F_y = (\hat{x}, \hat{y}) = \frac{\partial F(x, y)}{\partial y}.$$

Ahora definamos la segunda derivada funcional de varias variables.

Sea U un entorno del punto $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ en $\mathbf{R}^n$, y $f: U \to \mathbf{R}$, una función definida y continuamente diferenciable en U. Dícese que la función f es dos veces diferenciable en el punto $\hat{x}$ si existe una forma cuadrática Q tal que

$$f(\hat{x} + h) = f(\hat{x}) + f'(\hat{x})[h] + (1/2)\, Q(h) + r(h),$$

donde $r(h) = o(|h|^2)$.

La forma cuadrática es definida por la matriz simétrica integrada por las derivadas parciales $\left(\frac{\partial^2 f(\hat{x})}{\partial x_i \partial x_j}\right)$, $i, j = 1, \ldots, n$.

Pasemos al caso de dimensión infinita. Sean X, Y ciertos espacios normalizados, y $\mathfrak{U} \subset X$, un conjunto abierto. Si la aplicación $f: \mathfrak{U} \to Y$ es diferenciable en cada punto $x \subset \mathfrak{U}$, entonces será definida la aplicación $x \to f'(x)$ del conjunto $\mathfrak{U}$ en el espacio $\mathcal{L}(X, Y)$. Dado que $\mathcal{L}(X, Y)$ también es un espacio normalizado, se puede plantear la cuestión acerca de la existencia de la *segunda* derivada

$$f''(x) = (f')'(x) \in \mathcal{L}(X, \mathcal{L}(X, Y)).$$

Para $h_1 \subset X f''(x)[h_1] \in \mathcal{L}(X, Y)$ tomamos $h_2 \in X$, entonces será definida $f''(x)\ [h_1, h_2] = f''(x)[h_1][h_2]$. Por consiguiente, será definida la aplicación $f''(x): X \times X \to Y$ lineal en cada argumento. Análogamente se definen las derivadas de órdenes superiores.

Teorema (de derivadas mixtas). *Si para la aplicación* $f: \mathfrak{U} \to Y$ *existe la segunda derivada* $f''(\hat{x})$, *entonces para todos* $h_1, h_2 \in X$

$$f''(\hat{x})[h_1, h_2] = f''(\hat{x})[h_2, h_1]$$

(ATF, pág. 156).

Ejercicios

1. Citar un ejemplo de aplicación continua que no tenga (en el punto fijo) derivada en ninguna dirección.
2. Citar un ejemplo de aplicación que tenga derivada en cualquier dirección, pero que no tenga variación de Lagrange.

3. Citar un ejemplo de aplicación que tenga variación de Lagrange, pero que no tenga derivada de Gateaux.
4. Citar un ejemplo de aplicación que tenga derivada de Gateaux, pero que no tenga derivada de Fréchet.
5. Citar un ejemplo de aplicación que tenga derivada de Fréchet, pero que no sea estrictamente diferenciable.

1.4. Principales teoremas del cálculo diferencial en espacios normalizados

Citemos los teoremas que se usan más a menudo para resolver problemas extremales.

1.4.1. Teorema de superposición. *Sean X, Y, Z ciertos espacios normalizados; $\mathcal{U}$, el entorno del punto $\hat{x}$ en X; $\mathcal{V}$, el entorno del punto $\hat{y}$ en Y; $\varphi(\hat{x}) = \hat{y}$, $\psi : \mathcal{U} \to Z$, $f = \psi \circ \varphi : \mathcal{U} \to Z$, la superposición de las aplicaciones φ y ψ.*

Entonces, si ψ es diferenciable según Fréchet en el punto $\hat{y}$, y φ es diferenciable según Fréchet en el punto $\hat{x}$ (si es diferenciable según Gateaux, entonces tiene variación de Lagrange), f posee igual propiedad que φ en el punto $\hat{x}$ y, en este caso, respectivamente,

$$f'(\hat{x}) = \psi'(\hat{y}) \circ \varphi'(\hat{x}),$$

$$f'_G(\hat{x}) = \psi'(\hat{y}) \circ \varphi'_G(\hat{x}).$$

$$\delta f(\hat{x}, h) = \psi'(\hat{y})\,[\delta\varphi(\hat{x}, h)]\ \forall h \in X.$$

Si ψ es estrictamente diferenciable en $\hat{y}$, y φ es estrictamente diferenciable en $\hat{x}$, entonces f será estrictamente diferenciable en $\hat{x}$.

El teorema de superposición no tiene, generalmente, lugar si ψ es diferenciable tan sólo según Gateaux.

$\triangleleft$ Analicemos detalladamente dos casos extremales: variación de Lagrange y diferenciabilidad estricta.

A) Según la definición de la derivada de Fréchet,

$$\psi(y) = \psi(\hat{y}) + \psi'(\hat{y})\,[y - \hat{y}] + o(y - \hat{y}).$$

Por consiguiente,

$$\lim_{\lambda \to 0} \frac{f(\hat{x} + \lambda h) - f(\hat{x})}{\lambda} = \lim_{\lambda \to 0} \frac{\psi(\varphi(\hat{x} - \lambda h) - \psi(\hat{y})}{\lambda} =$$

$$= \lim_{\lambda \to 0} \frac{\psi'(\hat{y})\,[\varphi(\hat{x} - \lambda h) - \hat{y}] + o(\varphi(\hat{x} + \lambda h) - \hat{y})}{\lambda} =$$

$$= \psi'(\hat{y}) \lim_{\lambda \to 0} \left[\frac{\varphi(\hat{x} + \lambda h) - \hat{y}}{\lambda}\right] + \lim_{\lambda \to 0} \frac{o(\varphi(\hat{x} + \lambda h) - \hat{y})}{\lambda} =$$

$$= \psi'(\hat{y})\,[\varphi'(\hat{x}, h)] + \lim_{\lambda \to 0} \frac{o(\varphi(\hat{x} + \lambda h) - \hat{y})}{\|\varphi(x + \lambda h) - \hat{y}\|} \times$$

$$\times \lim_{\lambda\to 0}\left\|\frac{\varphi(\hat{x}+\lambda h)-\hat{y}}{\lambda}\right\| = \psi'(\hat{y})[\varphi'(\hat{x}, h)] +$$

$$+ 0\cdot\|\varphi'(\hat{x}, h)\| = \varphi'(\hat{y})[\varphi'(\hat{x}, h)],$$

lo cual precisamente demuestra la fórmula de variación de Lagrange de la superposición de las aplicaciones.

B) Como $\varphi \in SD(\hat{x})$, $\psi \in SD(\hat{y})$, para cualesquiera $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ se hallarán $\delta_1 > 0$, $\delta_2 > 0$ tales que de las desigualdades $\|x_i - \hat{x}\| < \delta_1$, $\|y_i - \hat{y}\| < \delta_2$, $i = 0, 1$, se deducirán las desigualdades

$$\|\varphi(x_1) - \varphi(x_2) - \varphi'(\hat{x})[x_1 - x_2]\| \leqslant \varepsilon_1 \|x_1 - x_2\|, \tag{1}$$

$$\|\psi(y_1) - \psi(y_2) - \psi'(y)[y_1 - y_2]\| \leqslant \varepsilon_2 \|y_1 - y_2\|. \tag{2}$$

Elijamos $\varepsilon_1 > 0$ y $\varepsilon_2 > 0$ para cualquier $\varepsilon > 0$, de tal modo que se cumpla la desigualdad

$$\varepsilon_1\varepsilon_2 + \varepsilon_2\|\varphi'(\hat{x})\| + \varepsilon_1\|\psi'(\hat{y})\| < \varepsilon.$$

Utilizando ε_1, ε_2 hallamos $\delta_1 > 0$, $\delta_2 > 0$, de tal modo que tengan lugar las relaciones (1) y (2) y suponiendo, por último, que

$$\delta = \min\{\delta_1, \delta_2/(\varepsilon_1 + \|\varphi'(\hat{x})\|)\}.$$

Y si ahora $\|x_i - \hat{x}\| < \delta$, $i = 0, 1$, en virtud de (1) tenemos

$$\|\varphi(x_1) - \varphi(x_2)\| \leqslant \varepsilon_1 \|x_1 - x_2\| + \|\varphi'(\hat{x}) \times$$

$$\times [x_1 - x_2]\| \leqslant \varepsilon_1 + \|\varphi'(\hat{x})\|)\|x_1x_2\|. \tag{3}$$

Suponiendo, en esta desigualdad, que $x_i = \hat{x}$, $i = 0,1$, consecutivamente, obtenemos

$$\|\varphi(x_i) - \varphi(\hat{x})\| = \|\varphi(x_i) - \hat{y}\| \leqslant (\varepsilon_1 + \|\varphi'(\hat{x})\|)\,\delta \leqslant \delta_2,$$

de tal modo que la relación (2) sea válida para $y_i = \varphi(x_i)$. Utilizando ahora las relaciones (1), (2) y (3), obtenemos

$$f(x_1) - f(x_2) - \psi'(\hat{y})[\varphi'(\hat{x})[x_1 - x_2]] \leqslant$$

$$\leqslant \|\psi(\varphi(x_1)) - \psi(\varphi(x_2)) - \psi'(\hat{y})[\varphi(x_1) - \varphi(x_2)]\| +$$

$$+ \|\psi'(\hat{y})[\varphi(x_1) - \varphi(x_2) - \psi'(\hat{x})[x_1 - x_2]]\| \leqslant$$

$$\leqslant \varepsilon_2\|\varphi(x_1) - \varphi(x_2)\| + \varphi'(\hat{y})\cdot\varphi(x_1) - \varphi(x_2) - \varphi'(\hat{x})[x_1 - x_2] \leqslant$$

$$\leqslant \varepsilon_2(\varepsilon_1 + \|\varphi'(\hat{x})\|)\|x_1 - x_2\| + \|\psi'(\hat{y})\|\varepsilon_1\|x_1 - x_2\| =$$

$$= (\varepsilon_2\varepsilon_1 + \varepsilon_2\|\varphi'(\hat{x})\| + \varepsilon_1\|\varphi'(\hat{y})\|)\|x_1 - x_2\| \leqslant \varepsilon\|x_1 - x_2\|,$$

esto precisamente significa que $f \in SD(\hat{x})$.

Suponiendo $x_2 = \hat{x}$, $x_1 = \hat{x} + h$ en nuestras consideraciones, obtenemos la demostración del teorema para el caso de diferenciabilidad de φ según Fréchet. La demostración del teorema para el caso de diferenciabilidad de φ según Gateaux se obtiene analizando la fórmula ya demostrada de la variación de Lagrange sobre la superposición de las aplicaciones.

1.4.2. Teorema del valor medio. Se sabe bien que para las funciones numéricas de una variable es válido el teorema de Lagrange también llamado **teorema del valor medio o fórmula de incrementos finitos:** *si la función* $f\colon [a, b] \to \mathbf{R}$ *es continua en el segmento* $[a, b]$ *y diferenciable en el intervalo* (a, b), *existirá un punto* $c \in (a, b)$ *tal que*

$$f(b) - f(a) = f'(c)(b - a). \tag{1}$$

Es fácil convencerse de que la fórmula (1) también es válida para las funciones numéricas $f(x)$ cuyo argumento pertenece a un espacio normalizado aleatorio. En este caso

$$[a, b] = \{x \mid x = a + t(b - a), \quad 0 \leqslant t \leqslant 1\}.$$

Análogamente se define el intervalo (a, b), mientras que la diferenciabilidad se puede comprender en el sentido de Gateaux. Suponiendo que $\varphi(t) = f(a + t(b - a))$ reducimos la demostración al caso de una variable real.

Para las funciones vectoriales la fórmula (1) no tiene lugar.

Como vemos, en el análisis se usa principalmente no la propia fórmula (1), sino la estimación $|f(b) - f(a)| \leqslant M(b - a)$ que de ella se desprende, donde $M = \sup\limits_{x \in (a, b)} f'(x)$. Mostremos ahora, que en esta versión más débil, la afirmación se extiende ya al caso de los espacios normalizados aleatorios. Tradicionalmente esa afirmación lleva el nombre de «teorema del valor medio», aunque, por supuesto, ella debería llamarse «teorema de estimación del incremento finito».

Teorema (del valor medio). *Supongamos que* X, Y *son espacios normalizados y que el conjunto abierto* $U \subset X$ *contiene el segmento* $[a, b]$.

Si la aplicación $f\colon U \to Y$ *es diferenciable según Gateaux en cada punto* $x \in [a, b]$, *entonces*

$$\| f(b) - f(a) \| \leqslant \sup_{c \in (a, b)} \| f'_G(c) \| \cdot \| b - a \|.$$

◁ En virtud del corolario del teorema de Hahn — Banach (p. 7.1), para cualquier $y \in Y$ y, por consiguiente, también para $y = f(b) - f(a)$ se hallará un elemento $y^* \in Y^*$ tal, que $\| y^* \| = 1$ e $\langle y^*, y \rangle = \| y \|$, es decir, $\langle y^*, f(b) - f(a) \rangle = \| f(b) - f(a) \|$.

Denotemos $\varphi(t) = \langle y^*, f(a + t(b - a)) \rangle$.

Dado que y^* es una funcional lineal continua, y la aplicación f posee en cada punto $[a, b]$ una derivada de Gateaux, entonces, según el teorema de superposición,

$$\varphi'(t) = \langle y^*, \ f'_G(a + t(b - a))[b - a] \rangle \ \forall t \in [0, 1].$$

De la diferenciabilidad de la función φ se deduce su continuidad en $[0, 1]$ y, por lo tanto, a ella se le puede aplicar la fórmula de Lagrange (de incrementos finitos):

$$\varphi(1) - \varphi(0) = \psi'(\theta), \quad \theta \in (0, 1).$$

Por eso

$$\| f(b) - f(a) \| = \langle y^*, f(b) - f(a) = \varphi(1) - \varphi(0) =$$
$$= \langle y^*, f'_G(a - \theta(b - a))[b - a] \rangle \leqslant \| f'_G(a +$$
$$+ \theta(b - a))[b - a] \| \leqslant \| f'_G(a + \theta(b - a)) \| \cdot \| b - a \| \leqslant$$
$$\leqslant \sup_{c \in (a, b)} \| f'_G(c) \| \cdot \| b - a \| \cdot \ \triangleright$$

Citemos algunos corolarios del teorema del valor medio.

Corolario 1. *Sean cumplidas todas las condiciones del teorema del valor medio y* $\Lambda \in \mathcal{L}(X, Y)$. Entonces

$$\| f(b) - f(a) - \Lambda(b - a) \| \quad \sup_{c \in (a, b)} \| f'_G(c) - \Lambda \| \cdot \| b - a \|.$$

$\triangleleft$ Es preciso usar el teorema del valor medio para la aplicación $g(x) = f(x) - \Lambda x$. $\triangleright$

Corolario 2. *Supongamos que X e Y son espacios normalizados, que U es el entorno del punto $\hat{x}$ en X, y que la aplicación $f : U \to Y$ es diferenciable según Gateaux en cada punto $x \in U$. Si la aplicación $x \to f'_G(x)$ es continua en el punto $\hat{x}$, la aplicación f será estrictamente diferenciable en $\hat{x}$ (y, por consiguiente, también será diferenciable según Fréchet en ese mismo punto).*

$\triangleleft$ En virtud de la continuidad de la aplicación $x \to f'_G(x)$, para cualquier $\varepsilon > 0$, se hallará $\delta > 0$ tal que de la desigualdad $\| x - \hat{x} \| < \delta$ se deducirá la desigualdad

$$\| f'_G(x) - f'_G(\hat{x}) \| < \varepsilon.$$

En virtud de la convexidad de la esfera $B = \mathring{B}(\hat{x}, \delta)$, de la condición $x_1, x_2 \in B$ se deduce que $[x_1, x_2] \in B$. Por el corolario 1 del teorema del valor medio con $\Lambda = f'_G(x)$ obtenemos

$$\| f(x_1) - f(x_2) - f'_G(\hat{x})[x_1 - x_2] \| \leqslant$$
$$\leqslant \sup_{x \in (X_1, X_2)} \| f'_G(x) - f'_G(\hat{x}) \| \, x \, \| x_1 - x_2 \| \leqslant$$
$$\leqslant \varepsilon \| x_1 - x_2,$$

lo cual demuestra la estricta diferenciabilidad de la aplicación f en el punto $\hat{x}$. $\triangleright$

El corolario 2 nos indica que para verificar la diferenciabilidad de las funcionales concretas es suficiente demostrar la presencia de la derivada de Gateaux y comprobar su continuidad, lo cual ya garantiza su diferenciabilidad estricta (así como la existencia de la derivada de Fréchet).

1.4.3. Fórmula de Taylor. Teorema de diferencial total.

Fórmula de Taylor. *Si existe* $f^{(n)}(\hat{x})$, *entonces*

$$f(\hat{x}+h) = f(\hat{x}) + f'(x)[h] + (1/2) f''(x)[h, h] + \ldots$$
$$\ldots + (1/n!) f^{(n)}(\hat{x})[h, \ldots, h] + r(h),$$

donde $\| r(h) \| = o(\| h \|^n)$ con $h \to 0$ [ATF, pág. 159].

Teorema de diferencial total. *Supongamos que* X, Y, Z *son espacios lineales normalizados;* U, *el entorno en* $X \times Y$; *y* $F: U \to Z$, *una aplicación que tiene en cada punto* $(x, y) \in U$ *las derivadas parciales de Gateaux* $F_x(x, y)$ *y* $F_y(x, y)$.

Si las aplicaciones $(x, y) \to F_x(x, y)$ *y* $(x, y) \to F_y(x, y)$ *son continuas en el punto* $(\hat{x}, \hat{y}) \in U$ *de una topología operacional uniforme,* F *será estrictamente diferenciable en ese mismo punto y, en este caso,*

$$F'(\hat{x}, \hat{y})[\xi, \eta] = F_x(\hat{x}, \hat{y})[\xi] + F_y(\hat{x}, \hat{y})[\eta].$$

◁ En virtud de la continuidad de las aplicaciones $F_x(x, y)$ y $F_y(x, y)$ en el punto $(\hat{x}, \hat{y})$, para cualquier $\varepsilon > 0$ se puede elegir $\delta > 0$ tal que el entorno «rectangular» $V = \mathring{B}(\hat{x}, \delta) \times \mathring{B}(\hat{y}, \delta)$ del punto $(\hat{x}, \hat{y})$ esté comprendido en U y en él se cumplan las desigualdades

$$\| F_x(x, y) - F_x(\hat{x}, \hat{y}) \| < \varepsilon, \quad \| F_y(x, y) - F_y(\hat{x}, \hat{y}) \| < \varepsilon.$$

Ahora tenemos

$$\Delta \overset{\text{def}}{=} F(x_1, y_1) - F(x_2, y_2) - F_x(\hat{x}, \hat{y}) \times [x_1 - x_2] -$$
$$- F_y(\hat{x}, \hat{y})[y_1 - y_2] = F(x_1, y_2) - F(x_2, y_1) -$$
$$- F_x(\hat{x}, \hat{y})[x_1 - x_2] + F(x_2, y_1) - F(x_2, y_2) -$$
$$- F_y(\hat{x}, \hat{y})[y_1 - y_2].$$

Es fácil ver que si los puntos (x_1, y_2), (x_2, y_2) se hallan en V, el punto $(x_2, y_1) \in V$ y también ambos segmentos $[(x_1, y_1), (x_2, y_1)]$ y $[(x_2, y_1), (x_2, y_2)]$ se encontrarán en $V \in U$. Por eso las aplicaciones $x \to F(x, y_1)$ e $y \to F(x_2, y)$ son diferenciables según Gateaux: la primera tiene la derivada F_x en $[x_1, x_2]$, y la segunda F_y en $[y_1, y_2]$. Empleando el corolario 1 del teorema del valor medio, con arreglo a dichas aplicaciones, en virtud de (1) obtenemos

$$\|\Delta\| \leqslant \sup_{\xi\in(x_1x_2)} \|F_x(\xi, y_1) - F_x(\hat{x}, \hat{y})\| \times$$
$$\times \|x_1 - x_2\| + \sup_{\eta\in(y_1, y_2)} \|F_y(x_2, \eta) -$$
$$- F_y(\hat{x}, \hat{y})\| \cdot \|y_1 - y_2\| \leqslant \varepsilon \|x_1 - x_2\| + \varepsilon \|y_1 - y_2\|$$

para cualesquiera $(x_1, y_1) \in V$, $(x_2, y_2) \in V$, lo cual precisamente demuestra la diferenciabilidad estricta de la aplicación F. $\triangleright$

1.4.4. Teoremas de dimensión finita de una función inversa e implícita. Teorema de Lusternik. Teorema de espacio tangente.

Teorema (de dimensión finita de una función inversa). *Supongamos que* $\mathcal{U} \in \mathbf{R}^n$ *es el entorno del punto* $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$, *y* $F : \mathcal{U} \to \mathbf{R}^n$, *una aplicación de clase* $C^1(\mathcal{U})$, $F(\hat{x}) = y$. *Entonces, si el jacobiano de la aplicación* F *en el punto* $\hat{x}$ *es distinto de cero, existirán tales* $\varepsilon > 0$, $\delta > 0$ *y* $K > 0$ *que para cualquier* y *de la esfera* $|\hat{y} - y| < \delta$ *existirá un sólo punto* x *en la esfera* $|x - \hat{x}| < \varepsilon$, *tal que* $F(x) = y$, *y, además,* $x - x \leqslant K |y - \hat{y}|$.

Nota. El teorema de función inversa ha sido expuesto de la forma en que será utilizado ulteriormente. Por regla general suele demostrarse más a menudo, en particular, que la suavidad de la aplicación inversa es igual que la de la suavidad directa, a la fórmula $(F^{-1}(y))' = (F'(F^{-1}(y)))^{-1}$.

Corolario (teorema de dimensión finita de una función implícita). *Sea* $\mathcal{U} = \mathbf{R}^k \times \mathbf{R}^s$ *el entorno del punto* $(\hat{x}, \hat{y}) = (\hat{x}_1, \ldots, \hat{x}_k, \hat{y}_1, \ldots, \hat{y}_s)$, *y* $\Psi : \mathcal{U} \to \mathbf{R}^s$, $\Psi(\hat{x}, \hat{y}) = 0$, $\Psi_y(\hat{x}, \hat{y})$, *una matriz invertible. Entonces existirán* $K > 0$, $\delta > 0$ *y una aplicación* $\varphi : B(\hat{x}, \delta) \to \mathbf{R}^s$ *de clase* $C^1(B(\hat{x}, \delta))$ *tal que*

$$\Psi(x, \varphi(x)) = 0, \quad \varphi(\hat{x}) = y, \quad |\varphi(x) - \hat{y}| \leqslant K(x - \hat{x}).$$

$\triangleleft$ Supongamos que $n = k + s$, $z = (x, y)$ $(x \in \mathbf{R}^k,\ y \in \mathbf{R}^s)$, $F(z) = (x, \Psi(z))$. Entonces la función F en el punto $\hat{z} = (\hat{x}, \hat{y})$ satisfará los requisitos del teorema de función inversa, puesto que

$$F'(\hat{z}) = \begin{bmatrix} I & \Psi_x & (\hat{z}) \\ 0 & \Psi_y & (\hat{z}) \end{bmatrix}$$

es una matriz regular. Según el teorema de función inversa, existen tales $\delta > 0$, $\varepsilon > 0$ y $K > 0$ que si $|\xi - \hat{x}| + |\eta| < \delta$, se hallará el único par (x, y) para el cual $|x - \hat{x}| + |y - \hat{y}| < \varepsilon$ y

$$F(x, y) = (\xi, \eta) \Leftrightarrow x = \xi, \quad \psi(x, y) = \eta,$$

$$|x - \hat{x}| + |y - \hat{y}| \leqslant K(|\xi - \hat{x}| + |\eta|).$$

Suponiendo que $\eta = 0$, resulta que si $|x - \hat{x}| < \delta$, existirá el único $y = \varphi(x)$, $|y - \hat{y}| < \varepsilon$ para el cual

$$F(x, \varphi(x)) = (x, 0) \Leftrightarrow \psi(x, \varphi(x)) = 0,$$

$$|y - \hat{y}| \leqslant K|x - \hat{x}|. \ \triangleright$$

Nota. De la fórmula de la derivada de función inversa se deduce inmediatamente que

$$\varphi'(x) = -[F_y(x, \varphi(x))]^{-1}[F_x(x, \varphi(x))].$$

Teorema de Lusternik. *Sean X, Z ciertos espacios banachianos, $\mathcal{U} \in \mathcal{O}(x, x)$ $F : \mathcal{U} \to Z$. Si $F \in SD(\hat{x})$ y $F'(\hat{x})$ es un epimorfismo, existirá el entorno $U \subset \mathcal{U}$ del punto $\hat{x}$, el número $K > 0$ y la aplicación $\varphi : U \to X$ tales que*

$$F(x - \varphi(x)) = F(\hat{x}), \quad \| \varphi(x) \| \leqslant KF(x) - F(\hat{x}) \|.$$

$\triangleleft$ La demostración de ese teorema se basa en el método modificado de Newton.

A) Sin limitar la generalidad estimamos que $\hat{x} = 0$ y $F(\hat{x}) = 0$. Elegimos $\varepsilon > 0$ tan pequeño que $\mathring{B}(0, \varepsilon) \subset \mathcal{U}$ y

$$\| F(x') - F(x'') - F'(0)(x' - X'') \| \leqslant \frac{1}{2C} \| x' - x'' \| \qquad (1)$$

para $\| x' \| < \varepsilon_1$, $\| x'' \| < \varepsilon$ (lo cual es posible ya que $F \in SD(x)$), donde la constante $C > 1$ fue tomada del lema acerca del operador inverso derecho (p. 7.2) para el operador M que también es inverso derecho respecto a $F'(0)$. Supongamos que para $x \in U = \mathring{B}(0, \delta)$,

$$\xi_{n+1} = \xi_n - M(F(\xi_n')), \quad n \geqslant 0, \quad \xi_0 = x, \qquad (2)$$

donde δ es tan pequeño que $\| x \| + O \| F(x) \| < \varepsilon/2$ para $\| x \| < \delta$.

B) Demostremos, por inducción que $\| \xi_n \| < \varepsilon \ \forall_n \geqslant 0$. Es evidente que $\| \xi_0 \| = \| x \| < \varepsilon/2$. Para $n = 1$, de la expresión 2 y del lema acerca del operador inverso derecho obtenemos la esti mación

$$\| \xi_1 - x \| = \| MF(x) \| \leqslant O \| F(x) \|, \qquad (3)$$

de donde $\| \xi_1 \| < \varepsilon/2$.

Sea $\| \xi_i \| < \varepsilon$ para $i = 0, 1, \ldots, k$ $(k \geqslant 1)$. De aquí deducimos que $\| \xi_{k+1} \| < \varepsilon$. Para $i = 0, 1, \ldots, k$ de (2) tenemos

$$F'(0)(\xi_{i-1}) - \xi_i) + F(\xi_i) = 0, \qquad (4)$$

de donde

$$\|\xi_{i+1}-\xi_i\|^{(2)}\leqslant C\,\|F(\xi_i)\|^{(4)}=C\,\|F(\xi_i)-$$
$$-F(\xi_{i-1})-F'(0)(\xi_i-\xi_{i-1})\|^{(1)}\leqslant\frac{1}{2}\|\xi_i-\xi_{i-1}\|\Rightarrow$$
$$\Rightarrow\|\xi_{i+1}-\xi_i\|\leqslant 2^{-i}\|\xi_1-x\|<2^{-1-i}\varepsilon,$$
$$i=1,\ \ldots,\ k. \tag{5'}$$

De aquí, en virtud de la desigualdad del triángulo, obtenemos

$$\|\xi_{k+1}\|=\|\xi_{k+1}-\xi_k+\xi_k-\xi_{k-1}+\ldots+\xi_2-$$
$$-\xi_1+\xi_1\|\leqslant\|\xi_{k+1}-\xi_k\|+\|\xi_k-\xi_{k-1}\|+$$
$$+\ldots+\|\xi_2-\xi_1\|+\|\xi_1\|<\frac{\varepsilon}{2}\left(\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^k}\right)+\frac{\varepsilon}{2}<\varepsilon.$$

Así pues, hemos obtenido que $\|\xi_{k+1}\|<\varepsilon$, de donde por inducción resulta que $\|\xi_n\|<\varepsilon\forall^n\geqslant 0$.

C) De las desigualdades (5), (5′) se desprende que

$$\|\xi_{n+m}-\xi_n\ \ |\ \xi_{n+1}-\xi_n\|\left(1+\frac{1}{2}+\ldots+\frac{1}{2^{m-1}}\right)\leqslant$$
$$\leqslant 2\,\|\xi_{n-1}-\xi_n\|\leqslant\frac{2}{2^n}\|\xi_1-x\|\to 0 \text{ para } n\to\infty,$$

es decir, $\{\xi_n\}\geqslant 0$ es una sucesión fundamental que converge en virtud del carácter banachiano de X.

Denotemos $\psi(x)=\lim\limits_{n\to\infty}\xi_n$; entonces

$$\|\xi_n-x\|\leqslant\|\xi_n-\xi_{n-1}\|+\|\xi_{n-1}-e_{n-2}\|+\ldots$$
$$\ldots+\|\xi_2-\xi_1\|+\|\xi_1-x\|\leqslant\|\xi_1-x\|\left(\frac{1}{2^{n-1}}+\frac{1}{2^{n-2}}+\ldots+1\right)\leqslant$$
$$\leqslant 2\,\|\xi_1-x\|.$$

Al pasar al límite obtenemos

$$\|\psi(x)-x\|\leqslant 2\,\|\xi_1-x\|\overset{(3)}{\leqslant}2C\,\|F(x)\|=K\,\|F(x)\|,$$

además,

$$\|\psi(x)\|\leqslant\|x\|+2\,\|\xi_1-x\|\overset{(3)}{<}\varepsilon\ .$$

De aquí y de (1) se deduce que F es continua en el punto $\psi(x)$ y, por eso, de (4) obtenemos

$$F(\psi(x))=\lim_{n\to\infty}F(\xi_n)=-\lim_{n\to\infty}F'(0)(\xi_{n+1}-\xi_n)=0.$$

Sea X un espacio normalizado, y M, cierto subconjunto de ese espacio. El elemento $h\in X$ se llama *vector tangente (semitangente) unilateral al conjunto M en el punto $\hat{x}\in M$* si existan $\varepsilon>0$ y la aplicación $r:[0,\ \varepsilon]\to X$ tales que:

a) $x+th+r(t)\in M\ \forall t\in[0,\ \varepsilon]$;
b) $\|r(t)\|=o(t)$ cuando $t\to+0$.

El vector h se llama *tangente al conjunto M en el punto $\hat{x}$*, siempre que los vectores h y $-h$ sean tangentes unilaterales a M en $\hat{x}$. El conjunto de todos los vectores tangentes a M en el punto $\hat{x}$ se designa por $T_{\hat{x}}M$, y el conjunto de los vectores tangentes unilaterales, por $T^{+}_{\hat{x}}M$. Es evidente que tanto $T_{\hat{x}}M$ como $T^{+}_{\hat{x}}$ son conos. Si el conjunto $T_{\hat{x}}M$ es un subespacio en X, el mismo se llama *espacio tangente a M en el punto $\hat{x}$*.

En muchos casos, incluido el que representa interés para la teoría de problemas extremales, el conjunto de vectores tangentes se define con ayuda de este corolario del teorema de Lusternik.

Teorema (del espacio tangente). *Sean X, Z ciertos espacios banachianos, $\mathcal{U} \in \mathcal{O}(\hat{x}, X)$, $F : \mathcal{U} \to Z$, $F \in SD(\hat{x})$ y $F(\hat{x})$, cierto epimorfismo, $M = \{x \in X \mid F(x) = F(\hat{x})\}$.*

Entonces

$$T_{\hat{x}}M = \operatorname{Ker} F'(\hat{x}).$$

A) Sea $h \in T_{\hat{x}}M$, $r(\cdot)$ la aplicación de la definición del vector tangente. Dado que $F \in SD(\hat{x})$, para pequeños valores de a:

$$F(\hat{x}) = F(\hat{x} + ah + r(a)) = F(\hat{x}) + aF'(\hat{x})[h] + o(a).$$

De aquí $aF'(\hat{x})[h] + o(a) = 0$ y, por lo tanto, $F'(\hat{x})[h] = 0$, es decir, $T_{\hat{x}}M \subset \operatorname{Ker} F'(\hat{x})$.

B) Sea $h \in \operatorname{Ker} F'(\hat{x})$. Supongamos que $r(a) = \varphi(\hat{x} + ah)$, donde φ es una aplicación construida en el teorema de Lusternik. Entonces

$$F(\hat{x} + ah - r(a)) = F(\hat{x}),$$

$$\| r(a) \| = \| \varphi(\hat{x} + ah) \| \leqslant K \| F(\hat{x} + ah) - F(\hat{x}) \| = o(a),$$

es decir, $h \in T_{\hat{x}}M$. $\triangleright$

Ejercicios

1. Construir dos aplicaciones φ y ψ tales que φ sea diferenciable según Fréchet, y ψ diferenciable según Gateaux, pero que el teorema de superposición no tenga lugar: $\psi \circ \varphi$ no es diferenciable según Gateaux.

2. Citar un ejemplo de aplicación para la cual no sea válida la fórmula de incrementos finitos: $f(b) - f(a) \neq f'(c)(b - a)\ \forall c \in [a, b]$.

3. Demostrar que si X es un espacio normalizado, $\mathcal{U} \in \mathcal{O}(\hat{x}, X)$, $F : \mathcal{U} \to \mathbb{R}^n$, $F(x) = (f_1(x), \ldots, f_n(x))$, $F \in D(\hat{x})$, entonces $F^1(\hat{x})$ será un epimorfismo tan sólo y solamente cuando las funcionales lineales $f'_1(\hat{x}), \ldots, f'_n(\hat{x})$ sean linealmente independientes.

1.5. Elementos del análisis convexo

Llámase análisis convexo la parte de las matemáticas donde se estudian los objetos convexos: conjuntos, funciones y problemas extremales.

1.5.1. Definiciones. Sea X un espacio lineal. El conjunto $A \subset X$ se llamará *convexo* si para cualesquiera dos puntos x_1 y x_2 de A y para cualquier número α del segmento $[0, 1]$, el elemento $\alpha x_1 + (1 - \alpha) x_2$ pertenece a A.

Todo triángulo en el plano es convexo. Entre los cuadrángulos existen tanto convexos como no convexos. Las esferas en un espacio normalizado son conjuntos convexos. En particular, son convexas, cuando $p \geqslant 1$, las esferas unidades $Bl_p^n := \{x \in \mathbf{R}^n \mid \sum_{i=1}^{n} |x_i|^p \leqslant 1\}$ de los espacios normalizados l_p^n.

También son convexos los semiespacios (es decir, los conjuntos de tipo $\{x \mid \langle x, \xi \rangle \leqslant \beta\}$ y sus intersecciones. La intersección de un número finito de semiespacios en $\mathbf{R}^n$ se llama *poliedro convexo*.

El conjunto K se llama *cono* si de la condición $x \in K$ se deduce que $\alpha x \in K$ para cualquier α positivo. Son *conos convexos* los representados por conjuntos convexos.

He aquí ejemplos de conos convexos: ángulo de abertura $\leqslant \pi$ en el plano; octante no negativo $\mathbf{R}_+^n$ en $\mathbf{R}^n$; conjunto de funciones no negativas en el espacio $C(T)$, donde T es un compacto; semiespacios acotados por hiperplanos que pasan por el cero en $\mathbf{R}^n$, así como sus intersecciones. La intersección de un número finito de semiespacios acotados por hiperplanos que atraviesan al cero se llama *ángulo poliedro con vértice en el cero.*

Supongamos que se ha dado la función $f: X \to \overline{\mathbf{R}}$ ($\overline{\mathbf{R}} := \mathbf{R} \cup \{\pm\infty\}$). Con cada una de tales funciones f se hallan relacionados dos conjuntos:

$\operatorname{dom} f := \{x \mid f(x) < +\infty$ *es un conjunto efectivo, y*

$\operatorname{epi} f := \{(\alpha, x) \in \mathbf{R} \times X \mid \alpha \geqslant f(x),\ x \in \operatorname{dom} f\}$, la *supergráfica de* f.

La función f se denomina *convexa* si la supergráfica de f es un conjunto convexo. La función $p: X \to \mathbf{R}^{\cdot}$ ($\mathbf{R}^{\cdot} = \mathbf{R} U \{+\infty\}$ se llama convexa homogénea o sublineal si la subgráfica de p es un cono convexo. También se puede decir así. La función p es sublineal si $p(\alpha x) = \alpha p(x)\ \forall \alpha > 0$, $x \ni X$ y $p(x_1 + x_2) \leqslant p(x_1) + p(x_2)\ \forall x_1, x_2 \in X$. La función f se llama propia si $f(x) > -\infty$ $\forall x$ ($\Leftrightarrow f: X \to \mathbf{R}^{\circ}$) y $\operatorname{dom} f \neq \varnothing$.

De la definición de conjunto convexo se deduce inmediatamente que la función propia será convexa cuando y sólo cuando se cumpla *la desigualdad de Jensen:*

$$f(\alpha x_1 - (1 - \alpha)x_2 \leqslant \alpha f(x_1) + (1 - \alpha) f(x_2)\ \forall x_1, x_2 \in X, \quad \alpha \in [0; 1].$$

Como ejemplos de funciones convexas de una variable son: los exponentes exp (ax), $a \in \mathbf{R}$, las funciones potenciales $|x|^p$, $p \geqslant 1$, y también x^β para $x > 0$ y $+\infty$ para $x \leqslant 0$ si $\beta < 0$. Como norma en cualquier espacio normalizado sirve el ejemplo de función sublineal.

Cualquier función afín en $\mathbf{R}^n$, es decir, la $a(x) := \sum_{i=1}^{n} \alpha_i x_i + \alpha$ es, por supuesto, convexa, y cualquier función lineal es sublineal. La función igual al máximo de cierta familia de funciones afines (lineales) es una función convexa (sublineal).

Los problemas extremales convexos se examinan en los párrafos 2 y 3.

Sea C cierto subconjunto finito X, es decir, $C := \{x_1, \ldots, x_n\}$. El elemento $x = \sum_{i=1}^{n} \alpha_i x_i$, $\alpha_i \geqslant 0$, $\sum_{i=1}^{n} \alpha_i = 1$ se llama *combinación convexa* de C, el elemento $\xi = \sum_{i=1}^{n} \beta_i x_i$, $\beta_i \geqslant 0$, *combinación cónica* de C. La totalidad de combinaciones convexas (cónicas) de los subconjuntos finitos del conjunto A se llama *envoltura convexa* (*cónica*) de A y se designa por co$A$ (cone A).

Se puede demostrar fácilmente que el conjunto co$A$ coincide con la intersección de todos los conjuntos que contienen A (esa propiedad se toma, a veces, por definición de co$A$), mientras que el conjunto Cone$A$ coincide con la intersección de todos los conos convexos que contienen A.

La envoltura convexa de un número finito de puntos se llama *poliedro convexo*, mientras que la envoltura cónica convexa de un número finito de puntos se denomina cono *finitamente generado*. Entre los poliedros convexos tienen tal estructura (en cierto grado) los simplex (envolturas convexas de un número finito de puntos afinamente independientes en $\mathbf{R}^n$), o sea, los triángulos en $\mathbf{R}^2$ y las pirámides triangulares en $\mathbf{R}^3$.

En la estructura de los conjuntos convexos desempeñan un papel especial los puntos extremos. Se dice que el punto x del conjunto convexo A es *extremo* si no existen puntos $x_1 \neq x_2$, $x_1, x_2 \in A$ y si los números λ, $0 < \lambda < 1$ son tales que $x = \lambda x_1 + (1 - \lambda x_2)$. Los puntos extremos son los vértices de los poliedros. Aquí tiene lugar el teorema de Minkowski: el conjunto compacto en R^n es la envoltura convexa de sus puntos extremos.

1.5.2. Operaciones con objetos convexos. Citemos algunas operaciones algebraicas y teóricas de conjunto que se realizan con los objetos convexos. Se puede demostrar elementalmente que dichas operaciones permiten transformar los objetos convexos en convexos.

Operaciones con las funciones. 1) Suma $(f_1 + f_2)(x) := f_1(x) + f_2(x)$; 2) convolución: $(f_1 \oplus f_2)(x) := \inf \{f_1(x_1) + f_2(x_2) \mid x_1 - x_2 = x\}$; 3) máximo: $(f_1 \vee f_2(x) := \max (f_1(x), f_2(x)$; 4) envoltura convexa del mínimo: $(f_1 \text{ co} \wedge f_2)(x) = \min \{\alpha f_1(x_1) + (1-\alpha) + (1-\alpha) f(x_2) \mid \alpha \in [0, 1], \alpha x_1 - (1-\alpha) x_2 = x\}$.

La tercera operación puede extenderse naturalmente a la familia aleatoria de funciones convexas.

Operaciones con los conjuntos: 1) suma: $A_1 - A_2 : - = \{x \mid x = x_1 - x_2,\ x_1 \in A_1,\ x_2 \in A_2$; 2) convolución — $A_1 \mid + \mid A_2 : = \bigcup (\alpha A_1 \cap (1 - \alpha) A_2),\ \alpha \in [0;\ 1]$; 3) intersección: $A_1 \cap A_2 = \{x \mid x \in A_1,\ x_2 \in A_2\}$; 4) envoltura convexa de una unión: $A_1 \operatorname{co} \bigcup A_2 = \operatorname{co} (A_1 \bigcup A_2)$.

La tercera operación puede aplicarse a cualquier familia de conjuntos convexos. Las operaciones 1), 3) y 4), tanto con las funciones como con los conjuntos, son absolutamente habituales, lo que no se puede decir acerca de las operaciones 2 (de convolución). Su sentido y significado serán aclarados cuando se trate de las relaciones duales.

Conforme a la idea general de este libro, en su primera parte nos limitaremos principalmente al caso de dimensión finita. En la segunda parte se hablará del análisis convexo en un espacio de dimensión infinita. En la primera parte X es $\mathbf{R}^n$.

Una de las propiedades más importantes de los objetos convexos es que ellos admiten la definición dual: en los espacios principal y «dual». En el caso de dimensión finita (cuando $X = \mathbf{R}^n$), el espacio dual (espacio de funcionales lineales) puede ser identificado con $\mathbf{R}^n$, facilitando asimismo nuestro problema.

Pasemos a la definición de los operadores duales.

Sea $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ cierta función. Llámase *transformación de Legendre — Young — Fenchel de la función* f la función

$$f^* (y) : = \sup_x (\langle x,\ y \rangle - f (x)).$$

De la definición de f^* se desprende que f^* es la cota superior de la familia de funciones afines, es decir, f^* es convexa.

La función $f^{**} (x) : \sup\limits_y (\langle x,\ y \rangle - f^* (y))$ se denomina *segunda conjugada* con f. De la definición de función conjugada se deduce la designaldad de Young:

$$\langle x,\ y \rangle \leqslant f(x) + f^*(y).$$

Es fácil calcular que la función afín $x \to a\ (x;\ \xi,\ \alpha) : = \sum\limits_{i=1}^{n} \xi_i x_i - \alpha$ será conjugada con la función «elemental» $e\ (y;\ \xi,\ \alpha)$, igual a α en el punto ξ y a $+\infty$ en los demás puntos y, al contrario, $e^*\ (^0;\ \xi,\ \alpha) = a\ (^0;\ \xi,\ \alpha)$.

Sea A un subconjunto no vacío en $\mathbf{R}^n$. Llámase *polar* del conjunto A el conjunto

$$A^0 : = \{y \in \mathbf{R}^n \mid \langle x,\ y \rangle \leqslant 1\ \forall x \in A\}.$$

Llámase *bipolar* el conjunto $A^{00} = \{x \in \mathbf{R}^n \mid \langle x,\ y \rangle \leqslant 1\ \ \forall y \in A^0$. De la definición de A^0 se deduce que aquí se trata de la intersección de una familia de semiespacios que contienen cero, es decir, de un conjunto convexo cerrado que contiene cero.

Es fácil comprender que la polar del segmento $[0; x]$ es el semiplano $\Pi_x := \{y \mid \langle x, y\rangle \leqslant 1$ y, viceversa, $\Pi_x^0 = [0, x]$. Notemos también que la polar de la esfera $B(0, r) := \{x \| x| \leqslant r\}$ es la esfera $B(0, r^{-1})$.

Sea K un cono en $\mathbf{R}^n$. Llámase *cono conjugado* con K el conjunto

$$K^* := \{y \in \mathbf{R}^n \mid \langle x, y\rangle \geqslant 0 \ \ \forall x \in K\}.$$

El cono $K^{**} := \{x \in \mathbf{R}^n \mid \langle x, y\rangle \geqslant 0 \ \ \forall y \in \mathbf{K}^*$ se llama *segundo cono conjugado* con K. De la definición de K^* se desprende que ese conjunto es la intersección de ciertos semiespacios cuyas fronteras son los hiperplanos que pasan por el cero. Por lo tanto, K^* es un cono convexo cerrado. Además, se ve con evidencia que $K^* = -K^0$.

Llámase *subdiferencial de la función sublineal* p el conjunto

$$\partial p := \{y \mid \langle x, y\rangle \leqslant p(x) \ \forall x\}.$$

Es fácil comprender que la subdiferencial de la norma euclídea $x \to |x|$ en $\mathbf{R}^n$ es la esfera $B(0, 1) = \{x \| x \leqslant 1\}$.

Llámase *subdiferencial de la función f (convexa propia) en el punto* $\hat{x}$ el siguiente conjunto:

$$\partial f(\hat{x}) := \{y \in \mathbf{R}^n \mid \langle x, -\hat{x}, y\rangle \leqslant f(x) - f(\hat{x}) \ \ \forall x\}.$$

De las definiciones se deduce inmediatamente que ∂p y $\partial f(\hat{x})$ son subconjuntos convexos en $\mathbf{R}^n$. Es fácil demostrar que ellos están cerrados.

Sea A un subconjunto no vacío en $\mathbf{R}^n$. La función

$$sA(y) := \sup\{\langle x, y\rangle \mid x \in A\}$$

se llama función *de apoyo* de A.

La subdiferencial de la función lineal $x \to \langle x, \xi\rangle$ coincide con $\{\xi\}$. La función de apoyo del elemento ξ es una función lineal de $\langle \circ, \xi\rangle$.

Después de cada definición se han dado ejemplos que podían parecer artificiales. En realidad, ellos entrañan gérmenes de aquellas relaciones de dualidad de las que vamos a tratar en los puntos que siguen.

Citemos dos definiciones más. Sea A un subconjunto no vacío en $\mathbf{R}^n$. La función

$$\mu A(x) := \inf\{\alpha > 0 \mid \alpha^{-1}x \in A\} \ (\inf \varnothing \ \ := + \infty)$$

se llama *función de Minkowski*, mientras que

$$\delta A(x) := \begin{cases} 0, & x \in A; \\ \infty, & x \in A \end{cases}$$

se den ina *función indicadora*. La función de Minkowski, correspondiente a la esfera Blp^n, es la norma en l_p^n.

$$\mu(Bl_p^n)(x)=\left(\sum_{i=1}^{n}\|x_i\|^p\right)^{1/p},\quad 1\leqslant p<\infty,$$

$$\mu(Bl_\infty^n(x))=\max_{1\leqslant i\leqslant n}|x_i|.$$

De la definición de μA (para A convexo) se deduce inmediatamente dicha función convexa homogénea (sublineal).

1.5.3. Teoremas de dualidad y compacidad. Enunciemos el sentido exacto de la declaración acerca de que los objetos convexos tienen doble descripción. Ese sentido es revelado por el teorema siguiente.

Teorema de dualidad. a) *La función* $f:\mathbf{R}^n\to\mathbf{R}$ *coincidirá con su segunda conjugada cuando y sólo cuando la misma sea convexa y cerrada* (es decir, *cuando* su supergráfica *sea cierto* conjunto convexo y cerrado).

b) *Sea* A *un conjunto no vacío en* $\mathbf{R}^n$. *El conjunto coincidirá con su bipolar cuando y sólo cuando* A *sea convexo, cerrado y contenga el cero.*

c) *El cono* $K\subset\mathbf{R}^n$ *coincidirá con su segundo conjugado, siempre que el mismo sea convexo y cerrado.*

d) *Sea* $p:\mathbf{R}^n\to\mathbf{R}$ *una función homogénea de primer grado con una subdiferencial no vacía. Para que se cumpla la igualdad* $s\partial p=p$ *es necesario y suficiente que* p *sea sublineal y cerrada.*

e) *Sea* A *un conjunto no vacío en* $\mathbf{R}^n$. *Para que se cumpla la relación* $\partial sA=A$ *es necesario y suficiente que* A *sea convexo y cerrado.*

La afirmación a) suele denominarse *teorema de Fenchel — Moreau*, y la afirmación b), *teorema de la bipolar*. Expresemos algunos corolarios del referido teorema, representándolos en forma de igualdades algebraicas o teóricas.

α) Sea f una función convexa cerrada propia. Entonces de a) se desprende que f, siendo, por un lado, la envoltura convexa de las funciones «elementales» asociadas $e\,(^0;\,x,\,f(x))$, $x\in\operatorname{dom}f$, por otro lado será la cota superior de una familia de funciones afines, o sea, de $\sup\{a(\cdot;\,y,\,f^*(y))\mid y\in\operatorname{dom}f^*\}$.

β) Sea A un subconjunto no vacío, convexo y cerrado en $\mathbf{R}^n$ y $0\in A$. Entonces $A=\bigcup_{x\in A}[0,\,x]=\bigcap_{y\in A^0}\Pi_y$, donde $\Pi_y=\{x\mid x,\,y\rangle\leqslant 1\}$.

γ) Sea p una función cerrada sublineal y $\partial p\neq\varnothing$. Entonces $p=\sup\ \{\langle\circ y\rangle/y\in\partial p\}$.

δ) Sea A un subconjunto no vacío convexo y cerrado. Entonces

$$A=\bigcup_{x\in A}\{x\}=\bigcap_{x\in\partial s\ A}\Pi(y,\,sA(y)),$$

$$\Pi(y,\alpha):=\{x\mid\langle x,\,y\rangle\leqslant\alpha\}.$$

Precisamente estas relaciones revelan la dualidad de definición de los objetos convexos. Reformulémoslas verbalmente.

La función convexa cerrada propia es la cota superior de las funciones afines que no la superan (Minkowski). La función sublineal cerrada con una subdiferencial no vacía es la cota superior de las funciones lineales que no la superan (Hermander). El subconjunto convexo cerrado es la intersección de los subespacios quo lo contienen (Minkowski).

En la segunda parte del libro (p. 8.1), el teorema de dualidad será demostrado totalmente en la variante de dimensión infinita. Todas las afirmaciones a) — d) tienen, en esencia, igual validez. Las afirmaciones b) — d) de la segunda parte del libro se deducen de a), o sea, del teorema de Fenchel — Moreau. Aquí demostraremos tan sólo la afirmación b), es decir, el teorema de la bipolar.

$\lhd$ De la definición de la bipolar se desprende que

$$A^{00} = \bigcap_{y \in A^0} \Pi_y, \ \Pi_y = \{x \mid \langle x, y\rangle \leqslant 1\}.$$

Está claro que Π_y es un subconjunto convexo cerrado que contiene cero. Por lo tanto, si $A = A^{00}$, entonces A será convexo, cerrado y contendrá cero. Ahora demostremos ese teorema en otro sentido. Si $x \in A$ e $y \in A^0$, de la definición resultará que $\langle x, y\rangle \leqslant 1$, es decir, $A \subset A^{00}$. Supongamos que tenemos el elemento $\hat{x} \in A^{00} \diagdown A$. Entonces (en virtud de la suposición de que A es convexo y cerrado), en el segundo teorema de separabilidad se hallará un elemento $\hat{y} \in \mathbf{R}^n$ tal que

$$\sup_{x \in A} \langle x, \hat{y}\rangle \leqslant 1, \quad \langle \hat{x}, \hat{y}\rangle \geqslant 1.$$

La primera desigualdad evidencia que $\hat{y} \in A^0$, lo cual significa que para $\hat{x} \in A^{00}$ debe cumplirse la desigualdad $\langle \hat{x}, \hat{y}\rangle \leqslant 1$, y eso contradice la segunda desigualdad anotada.

Señalemos el siguiente resultado.

Teorema de compacidad de la polar y la subdiferencial. a) *Sea A un entorno abierto de cero en $\mathbf{R}^n$. Entonces la polar A será compacta.* b) *Sea p una función continua sublineal en $\mathbf{R}^n$. Entonces la subdiferencial de p será un compacto convexo. Si p es una función sublineal cerrada, entonces ∂p será un conjunto convexo cerrado.*

Ambos resultados de compacidad admiten generalizaciones de dimensiones finitas. Sus demostraciones se exponen en el p. 8.1. En el caso de dimensión finita todo resulta muy sencillo.

$\lhd$ a) Ya hemos utilizado el hecho de que la polar es un conjunto convexo cerrado. Si la esfera $B(0, r) = \{x \parallel x \mid \leqslant r\}$ está contenida en A, para las polares tendremos un encaje inverso. Pero $(B, 0, r)^0 = B(0, r^{-1})$. Así, A^0 es un conjunto cerrado acotado en $\mathbf{R}^n$ es decir, un conjunto compacto.

b) Sea p una función sublineal continua. Siempre se puede estimar que $p \geqslant 0$ (sustrayendo, si no es así, la función lineal de la subdiferencial). Sea $M = \text{máx}\ \{p(x) \parallel x \mid = 1\}$. Entonces, en virtud de la homogeneidad de $p(x) \leqslant M \mid x \mid =: p_0(x)\ \forall x$ y, por consiguiente, $\partial p \subset \partial p_0(x) = B(0. M)$.

El carácter cerrado de ∂p es evidente, por lo tanto (en caso de la continuidad de p), ∂p es un conjunto acotado cerrado, o sea, un conjunto compacto.

1.5.4. Cálculo convexo. En el p. 1.5.2 fueron introducidas las operaciones con objetos convexos (sumas, convoluciones, etc.) y luego los operadores de dualidad (conjugación de funciones y conos, de la polar, de ∂ y, de s). El cálculo convexo constituye un conjunto de fórmulas de este tipo. Supongamos que el objeto convexo ha sido formado, digamos, por dos elementos mediante cierta operación. Es necesario expresar el operador dual de dicho objeto mediante los operadores duales de las componentes. En este sentido el cálculo convexo es afín al cálculo diferencial.

Las fórmulas del cálculo convexo se dividen en dos clases. Unas son válidas siempre y entonces escribimos $=$, y otras son vigentes al observarse determinadas condiciones. Entonces escribimos $\cong$.

A continuación se expone la tabla de las fórmulas principales.

I. Cálculo de la transformación de Legendre — Young — Fenchel.

1. $(f_1 \oplus f_2)^* = f_1^* + f_2^*$.
2. $(f_1 + f_2)^* \cong f_1^* \oplus f_2^*$.
3. $(f_1 \text{ co} \bigwedge f_2)^* = f_1^* \bigvee f_2^*$.
4. $(f_1 \bigvee f_2)^* \cong f_1^* \text{ co} \bigwedge f_2^*$.

II. Cálculo subdiferencial:

1. $\partial (p_1 + p_2) \cong \partial p_1 + \partial p_2$.
2. $\partial (p_1 \bigvee p_2) \cong \partial p_1 \text{ co} \bigcup \partial p_2$.

III. Cálculo de las polares:

1. $(A_1 + A_2)^0 = A_1^0 \,|\mp|\, A_2^0$.
2. $(A_1 |\mp| A_2) \cong A_1^0 + A_1^0$.
3. $(A_1 \bigcap A_2)^0 \cong A_1^0 \text{ co} \bigcup A_2^0$.
4. $(A_1 \text{ co} \bigcup A_2)^0 = A_1^0 \bigcap A_2^0$.

IV. Cálculo de las funciones de apoyo:

1. $s(A_1 + A_2) = sA_1 + sA_2$.
2. $s(A_1 \bigcap A_2) \cong sA_1 \oplus sA_2$ $(\cong sA_1 \text{ co} \bigwedge sA_2)$.
3. $s(A_1 \text{ co} \bigcup A_2) = sA_1 \bigvee sA_2$.

De las fórmulas I — IV y de las expresiones $K^* = -K^0$, $L^\perp = L^0$ resultan las relaciones explícitas para los anuladores y conos conjugados:

1. $(L_1 + L_2)^\perp = L_1^\perp \bigcap L_2^\perp$.
2. $(L_1 \bigcap L_2)^\perp \cong L_1^\perp + L_2^\perp$.

3. $(K_1 + K_2)^* = K_1^* \cap K_2^*$

4. $(K_1 \cap K_2)^* = K_1^* + K_2^*$.

Las fórmulas I.1, I.2, III.1 y III.2 revelan el sentido y la importancia de las operaciones de convolución para las funciones y conjuntos.

Exponemos la demostración de las dos fórmulas más importantes de la tabla presentada.

Fórmula de Moreau — Rocafellar. *Sean p_1 y p_2 ciertas funciones lineales, de las cuales p_1 es continua y p_2 cerrada. Entonces,*

$$\partial (p_1 + p_2) = \partial p_1 - \partial p_2.$$

$\lhd$ Si $y \in \partial p_1 + \partial p_2$, es decir, si $y = y_1 + y_2$, $y_i \in \partial p_1$, $i = 1, 2$, entonces $\langle y, y_i \rangle \leqslant p_i(x) \forall x$. Por consiguiente, $\langle x, y \rangle \leqslant (p_1 + p_2)(x) \, \forall x$, es decir, $y \in \partial (p_1 + p_2) \Rightarrow \partial p_1 + \partial p_2 \subset \partial (p_1 + p_2)$.

Supongamos que $\partial p_1 + \partial p_2 \neq \partial (p_1 + p_2)$ e $\hat{y} \in \partial (p_1 - p_2) \setminus (\partial p_1 + \partial p_2)$. En virtud del teorema de compacidad, ∂p_i es un compacto convexo, y ∂p_2, un conjunto convexo cerrado. Por lo tanto resulta que $\partial p_1 + \partial p_2$ es un conjunto convexo cerrado y, valiéndose del segundo teorema de separabilidad, se puede hallar un elemento $\hat{x}$ tal que

$$\sup_{y_i \in \partial p_i} \langle \hat{x}, y_1 + y_2 \rangle < \langle \hat{x}, \hat{y} \rangle.$$

Pero, por definición, $\sup \{ \langle \hat{x}, y_1 + y_2 \rangle \mid y_i \in \partial p_i, i = 1, 2\} \neq s\partial p_1 (\hat{x}) + s\partial p_2 (\hat{x}) =$ (según el teorema de dualidad d)) $= p_1(\hat{x}) + p_2(\hat{x})$. Resulta que $p_1(\hat{x}) + p_2(\hat{x}) < \langle \hat{x}, \hat{y} \rangle$. Eso contradice la hipótesis de que $\hat{y} \in \partial (p_1 - p_2)$. $\rhd$

Fórmula de Dubowizki — Miliutin. *Sean p_1 y p_2 ciertas funciones continuas sublineales. En este caso*

$$\partial (p_1 \vee p_2) = \operatorname{co} (\partial p_1 \cup \partial p_2).$$

$\lhd$ Si $y \in \operatorname{co} (\partial p_1 \cup \partial p_2)$, es decir, si $y = \alpha y_1 + (1 - \alpha) y_2$, $y_i \in \partial p_i$, $i = 1, 2$, $0 \leqslant \alpha \leqslant 1$, entonces $\langle x, y \rangle = \alpha \langle x, y_1 \rangle + (1 - \alpha) \langle x, y_2 \rangle \leqslant \alpha p_1(x) + (1 - \alpha) p_2(x) \leqslant (p_1 \vee p_2)(x)$, o sea, $\operatorname{co} (\partial p_1 \cup \partial p_2) \subset \partial (p_1 \vee p_2)$.

Supongamos que

$$\operatorname{co} (\partial p_1 \cup \partial p_2) \neq \partial (p_1 \vee p_2) \text{ e}$$

$$\hat{y} \in \partial (p_1 \vee p_2) \setminus \operatorname{co} (\partial p_1 \cup \partial p_2).$$

De la continuidad de p_i y del teorema de compacidad se desprende que ∂p_i son compactas. Pero en este caso la envoltura convexa de la asociación $\operatorname{co} (\partial p_1 \cup \partial p_2)$ también será compacta, o sea, será un conjunto cerrado. Por consiguiente, valiéndose del segundo teorema de separabilidad es posible hallar un elemento $\hat{x}$ tal que

$$a := \sup_{\alpha \in [0, 1],\, y_i \in \partial p_i,\ i=0,\ 1} \langle \hat{x},\ \alpha y_1 + (1-\alpha)\, y_2 \rangle < \langle \hat{x},\ \hat{y} \rangle.$$

De las definiciones se deduce inmediatamente que $a := \sup_{\alpha \in [0, 1]} (\alpha s\, \partial p_1 + (1-\alpha)\, s \partial p_2)(\hat{x}) =$ (según el teorema de dualidad d)) $= \sup_{\alpha \in [0, 1]} (\alpha p_1 (\hat{x}) + (1-\alpha)\, p_2)(\hat{x})) = \partial (p_1 \vee \vee p_2)(\hat{x})$, lo cual contradice la tolerancia de que $y \in \partial (p_1 \vee p_2)$. $\triangleright$

Para las funciones convexas existen fórmulas análogas

Si f_1 y f_2 son funciones convexas cerradas, cuando f_1 es continua en cierto punto $\bar{x}$, donde f_2 es finita, para cualquier punto $x \in \mathbf{R}^n$ obtenemos

$$\partial (f_1 + f_2)(x) = \partial f_1 (x) + \partial f_2 (x).$$

Si f_1 y f_2 son funciones convexas y continuas en el punto x, además, $f_1(x) = f_2(x)$, entonces

$$\partial (f_1 \vee f_2)(x) = \operatorname{co} (\partial f_1 (x) \cup \partial f_2 (x)).$$

Ambas fórmulas de Moreau—Rocafellar y de Dubowizki—Miliutin (para las funciones convexas) admiten generalizaciones de dimensión infinita. Veamos cómo se enuncia la generalización de esta última fórmula.

Teorema de depuración (para las subdiferenciales). *Supongamos que T es un compacto, y f: $T \times \mathbf{R}^n \to \mathbf{R}$, una función dotada de las siguientes propiedades:*

a) *$f(t, \cdot)$ es convexa y continua en $\mathbf{R}^n$ $\forall\, t \in T$;*

b) *$f(\cdot, x)$ es semicontinua por encima de $\forall \in \mathbf{R}^n$.*

Supongamos que $f(x) = \max_{t \in T} f(t, x)$, $y \in \partial f(\hat{x})$. Entonces existirán $r \leqslant n + 1$, $\tau_i \in T$ tales que $f(\tau_i, \hat{x}) = f(\hat{x})$,

$$y_i \in \partial_x f(\tau_i, \hat{x}) \text{ e } y \in \operatorname{co} \{y_1, \ldots, y_r\}.$$

La demostración de este teorema se ofrece en el punto 8.2.

Ejercicios

1. Demostrar que el cono K es convexo si y sólo si de la condición $x_1, x_2 \in K$ se desprende que $x_1 + x_2 \in K$.

2. Demostrar que no existe una función convexa acotada, determinada en toda la recta y diferente de la constante.

3. Demostrar que toda función convexa, siendo finita en toda la recta, es continua.

4. Demostrar que la función de Minkowski del conjunto convexo es convexa y homogénea.

5. Aclarar con qué valores paramétricos las referidas funciones resultan convexas:

a) $f(x) = ax^2 + bx + c$;

b) $f(x) = ae^{2x} + be^{x} + c$;

c) $f(x_1, x_2) = (|x_1|^p + |x_2|^p)^{1/p}$, $p > 0$;

d) $f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$.

6. Serán convexas las siguientes funciones:

a) $f(x) = x \ln x + (1 - x) \ln (1 - x)$, $x \in (0, 1)$;

b) $f(x) = \inf \{x_1^2 + x_2^2 \mid x_1 + x_2 = x\}$.

7. Hallar las funciones conjugadas de las siguientes funciones de una variable:

a) e^x; b) $ax^2 + bx + c$; c) $|x|^p/p$, $p > 0$;

d) $\delta\{0\}$ es la función indicadora del conjunto $A = \{0\}$;

e) $\delta[a, b]$; f) $f(x) = \begin{matrix} -\ln x, & x > 0, \\ +\infty, & x \leqslant 0. \end{matrix}$

8. Hallar las funciones conjugadas de las siguientes funciones de muchas variables:

a) $a_1x_1 + a_2x_2 + b$;

b) $a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$;

c) $e^{a_1x_1 + a_2x_2}$;

d) δB, donde $B = \{(x_1, x_2) \in \mathbf{R}^2 \mid x_1/a_1|^p + |x/a_2|^p \leqslant 1$, $p > 1$;

e) $f(x_1, \ldots, x_n) = \text{máx}\, \{x_1, \ldots, x_n^2\}$.

9. Hallar las segundas funciones conjugadas de las siguientes funciones:

a) $\sqrt{|x|}$; b) $(x^2 - 1)^2$; c) $\operatorname{sen} x$; d) $1/x^2$;

e) $|x| + |x - a|$; f) $||x| - 1|$.

10. Hallar las polares de los siguientes conjuntos en el plano:

a) $A = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$;

b) $A = \{(x_1, x_2) \mid x_1^2 - x_2^2 \leqslant 1\}$;

c) el triángulo con vértices en los puntos $(1, 0)$, $(-1/2, \pm\sqrt{3}/2)$;

d) $B_p = \{(x_1, x_2) \mid |x_1|^p + |x_2|^p \leqslant 1$, $p > 1$;

e) $A = \{(x_1, x_2) \mid x_1^2/a_1^2 + x_2^2/a_2^2 \leqslant 1\}$.

11. Calcular las subdiferenciales de las siguientes funciones convexas homogéneas de una variable:

a) $|x|$; b) $\text{máx}\, \{x, 0\}$; c) $\text{máx}\, \{-x, 0\}$.

12. Calcular las subdiferenciales de las siguientes funciones convexas homogéneas de muchas variables:

a) $|x| \overset{\text{def}}{=} \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$; b) $\underset{i=1, \ldots, n}{\text{máx}} |x_i|$;

c) $\underset{i=1, \ldots, n}{\text{máx}} x_i$, d) $\text{máx}\, \{0, \langle a, x\rangle\}$, $a \in \mathbf{R}^n$.

13. Hallar la subdiferencial de la norma (como una función convexa homogénea) en un espacio normalizado.

14. Hallar la subdiferencial $\partial f(\hat{x})$ de las siguientes funciones convexas:

a) $X = \mathbf{R}$, $f(x) = \text{máx}\, \{x^x, 1 - x\}$, $\hat{x} = 0$;

b) $X = C([0, 1])$, $f(x(\cdot)) = \underset{t \in [0,1]}{\text{máx}} |x(t)|$, $x(\hat{t}) = \operatorname{sen} 3\pi t$.

15. Hallar la función de Minkowski para un triángulo con vértices en los puntos

16. Hallar la función conjugada de la función de Minkowski.

17. Demostrar que cuando $f: \mathbf{R}^n \to \mathbf{R}$ y $f^*(x) = f(x)$ resulta que $f(x) = |x|^2/2$.

18. Mostrar, con un ejemplo, que la superposición de dos funciones convexas no es siempre convexa.

19. Mostrar, con un ejemplo, que la convolución de dos funciones convexas cerradas no siempre es una función convexa cerrada.

§ 2. Problemas suaves con igualdades y desigualdades. Problemas de programación convexa

En este párrafo se deducen las condiciones de extremo necesarias para los problemas suaves y convexos con acotaciones del tipo de igualdades y desigualdades. El material aquí usado se basa en los medios algebraicos y analíticos corrientes. Uno de los teoremas principales (p. 2.4) se demuestra directamente para un caso general, pero se entiende como el mismo puede ser expuesto en una variante de dimensión finita. Los principales resultados del párrafo se examinan otra vez en la segunda parte del libro (véanse los puntos 9.1 y 11.1).

2.1. Problemas no acotados

2.1.1. Extremos de las funciones de una variable. Sea $f : \mathbf{R} \to \mathbf{R}$ una función de una variable real. Analicemos el problema de búsqueda de los extremos de esa función:

$$f(x) \to \text{extr}.$$

Teorema 1 (de Fermi). *Si $\hat{x}$ es un punto del extremo local de la función f diferenciable en el punto $\hat{x}$, entonces*

$$f'(\hat{x}) = 0.$$

La demostración del teorema de Fermi figura ahora en los programas escolares y por eso aquí no la repetiremos. El teorema de Fermi proporciona la condición de primer orden que se necesita para la existencia de un extremo local. Formulemos las condiciones de primero y segundo órdenes.

Teorema 2. *Sea f una función dos veces diferenciable en el punto $\hat{x}$.*

Condiciones de extremo necesarias: *si $\hat{x}$ es un punto de mínimo (máximo) local de la función f, entonces*

$$f'(\hat{x}) = 0, \quad f''(\hat{x}) \geqslant 0 \quad (f''(\hat{x}) \leqslant 0.$$

Condiciones de extremo suficientes: *si*

$$f'(\hat{x}) = 0, \ f''(\hat{x}) > 0 \quad (f''(\hat{x}) < 0),$$

entonces x será el punto de mínimo (máximo) local de la función f.

Las demostraciones de los teoremas 1, 2 serán expuestas en el p. 2.1.2 para un caso más general.

En un caso unidimensional se puede ofrecer un análisis casi exhaustivo acerca de si el punto dado $\hat{x}$ es o no es un extremo local.

Teorema 3. *Sea f una función de una variable determinada en cierto intervalo que contiene el punto $\hat{x}$ y que es n veces diferenciable en este último.*

Condiciones de extremo necesarias: *si $\hat{x}$ es un punto de mínimo (máximo) local de la función f, entonces* $f'(\hat{x}) = \ldots = f^{n}(\hat{x}) = 0$ *o bien*

$$f'(\hat{x}) = \ldots = f^{(2m-1)}(\hat{x}) = 0, \qquad (1)$$
$$f^{(2m)}(\hat{x}) > 0 \quad (f^{(2m)}(\hat{x}) < 0)$$

para cierto $m \geqslant 1$, $2m \leqslant n$.

Condiciones de extremo suficientes: *si se cumple la condición (1), entonces $\hat{x}$ será el punto de mínimo (máximo) local de la función f.*

◁ Según la fórmula de Taylor, para una función n veces diferenciable en el punto $\hat{x}$ tenemos el siguiente desarrollo:

$$f(\hat{x}+x) = \sum_{k=0}^{n} \frac{f^{(k)}(\hat{x})}{k!} x^k + r(x),$$
$$\frac{r(x)}{x^n} \to 0 \quad \text{cuando } x \to 0.$$

La necesidad para $n = 1$ se deduce del teorema de Fermat. Sea en lo sucesivo $n > 1$. Entonces $f'(\hat{x}) = \ldots = f^{(n)}(\hat{x}) = 0$ o bien $f'(x) = \ldots = f^{(l-1)}(\hat{x}) = 0$, $f^{(l)}(\hat{x}) \neq 0$, $l \leqslant n$. Aquí es posible uno de los dos casos: l es impar o l es par. En el primer caso supongamos que $\varphi(\xi) = f(\hat{x} + \sqrt[l]{\xi})$, $\xi \in \mathbf{R}$. Entonces

$$\varphi(\xi) = f(\hat{x}) + \sum_{k=l}^{n} \frac{f^{(k)}(\hat{x})}{k!} \xi^{\frac{k}{l}} + r(\sqrt[l]{\xi})\,((r(\sqrt[l]{\xi}))\xi^{n/l} \to 0,$$

cuando $\xi \to 0$ será una función diferenciable en cero. Como $\hat{x} \in \in \operatorname{locmín} f$, por lo tanto, $0 \in \operatorname{locmín} \varphi$. De acuerdo con el teorema de Fermat, $\varphi'(0) = f^{(l)}(\hat{x})/l! = 0$. La contradicción muestra que l debe ser par: $l = 2m$. Por eso, de la fórmula de Taylor resulta que

$$f(\hat{x}+x) - f(\hat{x}) = \frac{f^{(2m)}(\hat{x})}{(2m)!} x^{2m} + r_1(x),$$
$$\frac{r_1(x)}{x^{2m}} \to 0 \quad \text{cuando } x \to 0.$$

Dado que $f^{(2m)}(\hat{x}) \neq 0$, entonces $f^{(2m)}(\hat{x}) > 0$ para $\hat{x} \in \operatorname{locmín} f$, y $f^{(2m)}(\hat{x}) < 0$ para $\hat{x} \in \operatorname{locmáx} f$.

Suficiencia. En vista de que $f'(\hat{x}) = \ldots = f^{(2m-1)}(\hat{x}) = 0$, $f^{(2m)}(\hat{x}) \neq 0$, según la fórmula de Taylor,

$$f(\hat{x}+x) - f(\hat{x}) = \frac{f^{(2m)}(\hat{x})}{(2m)!} x^{2m} + r_2(x),$$

$$\frac{r_2(x)}{x^{2m}} \to 0 \quad \text{para } x \to 0.$$

Por consiguiente, si $f^{(2m)}(\hat{x}) > 0$, entonces $f(\hat{x}+x) - f(\hat{x}) \geqslant \geqslant 0$ para x bastante pequeños, es decir, cuando $\hat{x} \in \operatorname{locmín} f$, si $f^{(2m)}(\hat{x}) < 0$, entonces $f(\hat{x}+x) - f(\hat{x}) \leqslant 0$ para x bastante pequeños, es decir, cuando $\hat{x} \in \operatorname{locmáx} f$. $\triangleright$

2.1.2. Extremos de las funciones de varias variables y de las funcionales. Sea f la aplicación del espacio normalizado X al conjunto de números reales $\mathbf{R}$, dotado de cierta suavidad, es decir, de determinadas propiedades de diferenciabilidad. Llámase *problema suave no acotado* el que presupone la búsqueda de los extremos de la siguiente función:

$$f(x) \to \text{extr}. \tag{Pr}$$

Teorema 1 (análogo al teorema de Fermat para los espacios normalizados). *Sea X un espacio normalizado y la funcional f, determinada en el entorno del punto $\hat{x}$, diferenciémosla según Fréchet (la misma posee variación según Lagrange) en el punto $\hat{x}$.*

En este caso, si $\hat{x} \in \operatorname{locextr} f$, tenemos

$$f'(\hat{x}) = 0 \quad (\delta f(\hat{x}, h) = 0 \quad \forall h \in X).$$

$\triangleleft$ Si $\hat{x} \in \operatorname{locextr} f$, entonces el punto cero será para el conjunto $\forall h \in X$ el extremo local de la función de una variable: $\lambda \to \varphi(\lambda; h) \overset{\text{def}}{=} \overset{\text{def}}{=} f(\hat{x} + \lambda h)$. Por eso $\varphi'_\lambda(0; h) = 0$ y, utilizando la definición de la variación según Lagrange, obtenemos $\delta f(\hat{x}, h) = 0$. Si la funcional f es diferenciable según Fréchet en el punto $\hat{x}$, entonces ella en este punto tendrá variación según Lagrange. Dado que $\hat{x} \in \in \operatorname{locextr} f$, de lo ya demostrado se deduce que la referida variación es igual a cero. De aquí $f'(\hat{x}) = 0$ en virtud de la definición de la diferenciabilidad según Fréchet (p. 1.3.2).

Del teorema 1 se desprende que si el punto $\hat{x}$ proporciona un extremo local a la función de varias variables $f\colon \mathbf{R}^n \to \mathbf{R}$, que es diferenciable en el punto $\hat{x}$, entonces todas las derivadas parciales de la función f en ese punto se transformarán en cero, es decir,

$$f'(\hat{x}) = 0 \Leftrightarrow \frac{\partial f(\hat{x})}{\partial x_1} = \ldots = \frac{\partial f(\hat{x})}{\partial x_n} = 0.$$

Teorema 2. *Sea X un espacio normalizado, $U \in \mathcal{O}(\hat{x}, X)$, $f: U \to \bar{\mathbb{R}}$, $f \in R^2(\hat{x})$.*

Condiciones de extremo necesarias: *si* $\hat{x} \in$ locmín (máx) f, *entonces* $f'(\hat{x}) = 0$, $f''(\hat{x})[x, x] \geqslant 0$ $(f''(\hat{x})[x, x] \leqslant 0)$ $\forall x \in X$.

Condiciones de extremo suficientes: *si* $f'(\hat{x}) = 0$ *y*

$$f''(\hat{x})[x, x] \geqslant \alpha \|x\|^2 \; (f''(\hat{x})[x, x] \leqslant -\alpha \|x\|^2) \; \forall x \in X \quad (1)$$

para cierto $\alpha > 0$, entonces $\hat{x} \in$ locmín (máx) f.

◁ Conforme a la fórmula de Taylor (p. 1.4.3),

$$f(\hat{x} + x) = f(\hat{x}) + f'(\hat{x})[x] + (1/2) f''(\hat{x})[x, x] + r(x),$$
$$\|r(x)\| = o(\|x\|^2).$$

Demostremos el teorema para el caso de mínimo. El caso de máximo es análogo.

Necesidad. En vista de que $\hat{x} \in$ locmín f, en primer lugar, según el teorema de Fermat (p. 2.1.3), $f'(\hat{x}) = 0$, y en segundo lugar, $f(\hat{x} - \lambda x) - f(\hat{x}) \geqslant 0$ cuando λ son bastante pequeños. Por eso, en virtud de la fórmula de Taylor, $f(\hat{x} - \lambda x) - f(\hat{x}) = (\lambda^2/2) \times \times f''(\hat{x})[x, x] + r(\lambda x) \geqslant 0$ cuando λ son pequeños. De aquí $f''(\hat{x})[x, x] \geqslant 0$ $\forall x \in X$.

Suficiencia. Dado que $f'(\hat{x}) = 0$, según la fórmula de Taylor, en virtud de la condición $f''(\hat{x})[x, x] \geqslant \alpha \|x\|^2$, tenemos $f(\hat{x} + x) - f(\hat{x}) = \frac{1}{2} f''(\hat{x}), [x, x] + r(x) \geqslant \frac{\alpha}{2} \|x\|^2 + r(x) \geqslant 0$ para x bastante pequeños. Por consiguiente, $\hat{x} \in$ locmín f. ▷

La definición no negativa de la segunda derivada para las funciones de n variables presupone la definición no negativa de la matriz $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$.

La condición (1) se llama *condición de positividad (negatividad)* estricta de la segunda derivada (en el sentido de Fréchet) de la funcional f.

Notemos que en los espacios de dimensión finita, la condición de definición positiva de la matriz $\frac{\partial^2 f(\hat{x})}{\partial x_i \partial x_j}$, es decir, la condición

$$\sum_{, j=1}^{n} \frac{\partial^2 f(\hat{x})}{\partial x_i \partial x_j} h_i h_j > 0 \quad \forall h = (h_1, \ldots, h_n) \in \mathbb{R}^n, \quad h \neq 0,$$

garantiza una positividad estricta de la segunda diferencial (y, por consiguiente, es una condición de mínimo suficiente en el punto estacionario). Pero en los espacios de dimensión infinita eso no es así [ATF, pág. 242].

Las definiciones positiva y negativa de la matriz se establecen mediante el criterio de Sylvester.

Teorema (criterio de Sylvester). *La matriz A es positivamente (negativamente) definida si y sólo si sus menores principales* $\det A_k$, *donde* $A_k = (a_{ij})_{i,j=1}^{k}$, $k = 1, \ldots, n$, son positivos ($(-1)^k \times \det A_k > 0$, $k = 1, \ldots, n$).

La demostración de esta afirmación se expone en [AGT, pág. 219].

2.2. Problema suave de dimensión finita con acotaciones del tipo de igualdades

2.2.1. Planteamiento del problema. Sean $f_i : \mathbf{R}^n \to \overline{\mathbf{R}}$, $i = 0, 1, \ldots, m$ las funciones de n variables que reflejan el espacio $\mathbf{R}^n$ en $\overline{\mathbf{R}}$. Llámase *problema extremal de dimensión finita con acotaciones del tipo de igualdades*, el siguiente problema en $\mathbf{R}^n$:

$$f_0(x) \to \text{extr}; \quad f_i(x) = 0, \ i = 1, \ldots, m. \qquad \text{(Pr)}$$

Seguidamente consideramos que todas las funciones f_i poseen determinada suavidad.

2.2.2. Regla de los factores de Lagrange. Teorema. *Supongamos que $\hat{x}$ es un punto de extremo local en el problema* (Pr) *y que las funciones f_i, $i = 0, 1, \ldots, m$, son continuamente diferenciables en el entorno del punto $\hat{x}$ (condición de suavidad). Entonces existirá un vector no nulo de los factores de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbf{R}^{m+1}$, tal que para la función lagrangiana del problema* (Pr)

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x)$$

se cumpla la condición de estacionaridad

$$\mathcal{L}_x(\hat{x}, \lambda) = 0 \iff \frac{\partial \mathcal{L}(\hat{x}, \lambda)}{\partial x_j} = 0,$$

$$j = 1, \ldots, n \iff \sum_{i=0}^{m} \lambda_i f_i'(\hat{x}) = 0.$$

◁ Demostrémoslo partiendo de lo contrario. Supongamos que no se cumple la condición de estacionaridad, es decir, los vectores $f_i'(\hat{x})$, $i = 0, 1, \ldots, m$ son linealmente independientes. Eso quiere decir que el rango de la matriz $A = \left(\frac{\partial f_i(\hat{x})}{\partial x_j}\right)_{\substack{i=0,\ldots,m \\ j=1,\ldots,n}}$ es igual a $m + 1$. Entonces, en virtud del teorema de rango matricial, existe una matriz M de orden $(m + 1)(m + 1)$ con un determinante distinto de cero. Supongamos, para mayor precisión, que dicha matriz es la integrada de las primeras columnas de la matriz Λ:

$$\det \begin{bmatrix} \frac{\partial f_0(\hat{x})}{\partial x_1} & \ldots & \frac{\partial f_0(\hat{x})}{\partial x_{m+1}} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m(\hat{x})}{\partial x_1} & \ldots & \frac{\partial f_m(\hat{x})}{\partial x_{m+1}} \end{bmatrix} = \det M \neq 0.$$

Estimamos, sin afectar la generalidad, que $f_0(\hat{x}) = 0$. Efectivamente, si $f_0(\hat{x}) \neq 0$, entonces hay que someter a consideración la función $\tilde{f}_0(x) = f_0(x) - f(\hat{x})$. Supongamos que para el vector $\bar{x} = (x_1, \ldots, x_{m+1})$: $F(x) = (F_0(x), \ldots, F_m(\bar{x})) = (f_0(x, \hat{x}_{m+2}, \ldots, \hat{x}_n), \ldots f_m(\bar{x}, \hat{x}_{m+2}, \ldots, \hat{x}_n))$; F representa cierto entorno del punto $\hat{\bar{x}} = (\hat{x}_1, \ldots, x_{m+1}) \in \mathbf{R}^{m+1}$ en $\mathbf{R}^{m+1}$ y es (en virtud de las condiciones de suavidad del teorema) una aplicación continuamente diferenciable de ese entorno, $F(\hat{\bar{x}}) = 0$. Además,

$$\det \left(\frac{\partial F_i(\hat{\bar{x}})}{\partial x_j} \right)_{\substack{i=0,1,\ldots,m \\ j=1,\ldots,m+1}} = \det M \neq 0.$$

Conforme al teorema de función inversa, en los espacios de dimensión finita (p. 1.4.4) existe una aplicación inversa F^{-1} de cierto entorno de cero, tal que $|F^{-1}(y) - \bar{x}| \leqslant K |y|$ con cierta constante $K > 0$. En particular, para ε bastante pequeño en módulo habrá un vector $\bar{x}(\varepsilon) = F^{-1}(\varepsilon, 0, \ldots, 0)$ tal que $F(\bar{x}(\varepsilon)) = (\varepsilon, 0, \ldots, 0)$, es decir,

$$f_0(x(\varepsilon)) = \varepsilon, \ f_i(x(\varepsilon)) = 0, \ i = 1, \ldots, m,$$

donde $x(\varepsilon) = (x_1(\varepsilon), \ldots, x_{m+1}(\varepsilon), \hat{x}_{m+2}, \ldots, \hat{x}_n)$ y además

$$|\bar{x}(\varepsilon) - \hat{\bar{x}}| \leqslant K|\varepsilon| \Leftrightarrow |x(\varepsilon) - \hat{x}| \leqslant K|\varepsilon|. \qquad (2)$$

De las relaciones 1 y 2 se deduce que el vector $\hat{x}$ no proporciona extremo al problema, ya que cerca de él existen los vectores admisibles $x(\varepsilon)$, en los cuales la funcional f_0 adopta valores tanto máximos como mínimos de $f_0(\hat{x})$ (recordemos que $f_0(\hat{x}) = 0$). Por lo tanto, hemos obtenido una contradicción con arreglo a que $\hat{x} \in \in$ locextr Pr. Así pues, nuestra suposición (contraria) es incorrecta y, por consiguiente, el teorema está demostrado. $\triangleright$

Nota 1. De la relación (1) se desprende que si los vectores $f_1'(\hat{x}), \ldots, f_m'(\hat{x})$ son linealmente independientes, entonces $\lambda_0 \neq 0$.

Nota 2. Hablando en general, se puede prescindir del tipo de extremo al recurrir a la regla de los factores de Lagrange para problemas con acotaciones del tipo de igualdades y, al convencerse de que $\lambda_0 \neq 0$, adoptar λ_0 por cualquier constante distinta de cero. Para los problemas donde participan desigualdades e inclusiones, el signo λ_0 tiene validez.

Formulemos en ese mismo punto las condiciones de extremo necesarias para los problemas con igualdades y desigualdades. Sean f_i: $\mathbf{R}^n \to \mathbf{R}$, $i = 0, 1, \ldots, m$ ciertas funciones de n variables. Llámase *problema extremal de dimensión finita con acotaciones del tipo de igualdades y desigualdades*, el siguiente problema en $\mathbf{R}^n$:

$$f_0(x) \to \inf; \quad f_i(x) \leqslant 0, \quad i = 1, \ldots, m',$$
$$f_i(x) = 0, \quad i = m' + 1, \ldots, m. \tag{Pr}$$

Teorema. *Supongamos que $\hat{x}$ es un punto de mínimo local en el problema* (Pr) *y que las funciones f_i, $i = 0, 1, \ldots, m$, son continuamente diferenciables en cierto entorno del punto $\hat{x}$ (condición de suavidad). Entonces existirá un vector no nulo de los factores de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbf{R}^{m+1}$ tal que para la función lagrangiana del problema* (Pr)

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x)$$

se cumplan las condiciones:

a) *de estacionaridad*

$$\mathcal{L}_x(\hat{x}, \lambda) = 0 \Leftrightarrow \sum_{i=0}^{m} \lambda_i f'(\hat{x}) = 0 \Leftrightarrow \frac{\partial \mathcal{L}(\hat{x}, \lambda)}{\partial x_j} = 0, \quad j = 1, \ldots, n;$$

b) *de no rigidez complementaria*

$$\lambda_i f_i(\hat{x}) = 0, \quad i = 1, \ldots, m';$$

c) *de no negatividad*

$$\lambda_i \geqslant 0, \quad i = 0, 1, \ldots, m'.$$

La demostración de dicho problema se ofrecerá en el p. 2.4 con arreglo a un caso más general.

2.3. Problemas de programación convexa

2.3.1. Problemas sin acotaciones. *Llámase problema convexo sin acotaciones* el problema siguiente:

$$f(x) \to \inf. \tag{Pr}$$

Aquí f: $X \to \overline{\mathbf{R}}$ es una función convexa que aplica cierto espacio lineal X a una recta ampliada.

Teorema (análogo al teorema de Fermat). *Para que el punto x proporcione al problema* (Pr) *un mínimo absoluto es necesario y sufi-*

ciente que se cumpla la relación

$$0 \in \partial f(\hat{x}).$$

$\lhd$ $x \in \operatorname{absmín} \mathrm{Pr} \Leftrightarrow f(x) - f(\hat{x}) \geqslant 0 = \langle 0, x - \hat{x} \rangle \Leftrightarrow 0 \in \partial f(\hat{x})$.

2.3.2. Planteamiento del problema de programación convexa. *Llámase problema de programación convexa* (o *problema convexo*) el siguiente problema extremal:

$$f_0(x) \to \inf; \; f_i(x) \leqslant 0, \; i = 1, \ldots, m, \; x \in A. \qquad \text{(Pr)}$$

Aquí f_i: $X \to \overline{\mathbf{R}}$, $i = 0, 1, \ldots, m$, son funciones convexas (funcionales) que aplican cierto espacio lineal X (no obligatoriamente normalizado) a una recta ampliada; A es un subconjunto convexo en X.

En vista de que de la convexidad de la función f no se desprende, hablando en general, la convexidad de $-f$, resulta que (Pr) no es un problema de maximización, sino de minimización.

El punto x se llama *admisible* en (Pr) si $x \in A$ y $f_i(x) \leqslant 0$, $i = 1, \ldots, m$.

Lema. *Sea X un espacio normalizado. En un problema convexo el mínimo local también es global.*

$\lhd$ Sea $\hat{x} \in \operatorname{locmín} \mathrm{Pr}$. Eso significa que existe un entorno $\mathcal{U}$ del punto $\hat{x}$, tal que $-\infty < f_0(\hat{x}) \leqslant f_0(x)$ para cualquier punto admisible $x \in \mathcal{U}$. Tomemos el punto admisible aleatorio x. Entonces, para $\alpha > 0$ bastante pequeño, el vector $\bar{x} = (1 - \alpha)\hat{x} + \alpha x \in \mathcal{U}$ precisamente será admisible. Por consiguiente, de la desigualdad de Jensen (p. 1.5.1) $f_0(\hat{x}) \leqslant f_0(\bar{x}) \leqslant (1 - \alpha) f_0(\hat{x}) + \alpha f(x)$, de donde $f_0(\hat{x}) \leqslant f_0(x)$. $\rhd$

Por eso ulteriormente, en los problemas convexos, al decir «mínimo» tendremos en cuenta el mínimo absoluto.

2.3.3. Teorema de Kuhn—Tacker. Sea X un espacio lineal; f_i: $X \to \overline{\mathbf{R}}$, $i = 0, 1, \ldots, m$, *ciertas funciones convexas en* X; *y* A, *un subconjunto convexo en* X.

1. *Entonces, si* $\hat{x}$ *es la solución del problema de programación convexa, se hallará un vector no nulo de los factores de Lagrange* $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, *tal que para la función lagrangiana* $\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x)$ *se cumplan:*

a) *el principio de mínimo para la función de Lagrange*

$$\min_{x \in A} \mathcal{L}(x, \lambda) = \mathcal{L}(\hat{x}, \lambda);$$

b) *la condición de no rigidez complementaria*

$$\lambda_i f_i(\hat{x}) = 0, \quad i = 1, \ldots, m;$$

c) *la condición de no negatividad*

$$\lambda_i \geqslant 0, \ i = 0, 1, \ldots, m.$$

2. *Si* $\lambda_0 \neq 0$, *entonces las condiciones* a) — c) *serán suficientes para que el punto admisible* $\hat{x}$ *sea la solución del problema.*

3. *Para que* $\lambda_0 \neq 0$ *es suficiente que se cumpla la condición de Slater, es decir, la condición de existencia del punto* $\bar{x} \in A$, *para el cual* $f_i(\bar{x}) < 0, \ i = 1, \ldots, m$.

$\triangleleft$ Sea $\hat{x}$ la solución del problema. Estimemos, sin afectar la generalidad, que $f_0(\hat{x}) = 0$, de lo contrario introduzcamos una nueva función $\tilde{f}_0(x) = f_0(x) - f_0(\hat{x})$. Supongamos que

$$B = \{b = (b_0, b_1, \ldots, b_m) \in \mathbf{R}^{m+1} \mid \exists x \in A: f_i(x) \leqslant b_i, \quad i = 0, 1, \ldots, m\}. \tag{1}$$

A) B es un conjunto convexo no vacío. Efectivamente, $\mathbf{R}_+^{m+1} \in B$, es decir, todo vector de componentes no negativas pertenece a B, ya que en (1) se puede suponer que $x = \hat{x}$. Demostremos su convexidad. Supongamos que b y b' pertenecen a B, $0 \leqslant \alpha \leqslant 1$, y que x y x' son tales elementos de A que (según la relación 1) $f_i(x) \leqslant b_i$, $f_i(x') \leqslant b_i'$, $i = 0, 1, \ldots, m$. Supongamos también que $x_\alpha = \alpha x + (1 - \alpha) x'$. Entonces, $x_\alpha \in A$, ya que A es convexo y, en virtud de la convexidad de las funciones f_i, $i = 0, 1, \ldots, m$, tenemos

$$f_i(x_\alpha) = f_i(\alpha x + (1 - \alpha) x') \leqslant \alpha f_i(x) + (1 - \alpha) f_i(x') \leqslant \leqslant \alpha b_i + (1 - \alpha) b_i',$$

es decir, el punto $\alpha b + (1 - \alpha) b' \in B$.

Denotemos $C = \{c = (c_0, 0, \ldots, 0) \in \mathbf{R}^{m+1} \mid c_0 < 0\}$.

B) C es un conjunto convexo no vacío, y $C \cap B = \varnothing$. Efectivamente, si existiera un punto $c = (c_0, 0, \ldots, 0)$, $c_0 < 0$ perteneciente a B, en virtud de la definición (1), de aquí se deduciría que existe un elemento $\tilde{x} \in A$ para el cual se cumplen las desigualdades $f_0(\tilde{x}) \leqslant \leqslant c_0 < 0$, $f_i(\tilde{x}) \leqslant 0$, $i = 1, \ldots, m$. Pero de estas últimas se deduce que $\hat{x}$ no es la solución del problema. Por eso $C \cap B = \varnothing$.

Según el primer teorema de separabilidad, los conjuntos B y C pueden ser separados en el caso de dimensión finita (p. 1.2.2), es decir, existe un vector $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \neq 0$ para el cual

$$\inf_{b \in B} \sum_0^m \lambda_i b_i \geqslant \sup_{c \in C} \sum_0^m \lambda_i c_i. \tag{2}$$

En vista de que $0 \in B$, por lo tanto, según la expresión (2), $0 \geqslant \geqslant \sup\limits_{c \in C} \sum\limits_0^m \lambda_i C_i = \sup\limits_{c \in C} \lambda_0 C_0$. De aquí $\lambda_0 \geqslant 0$ y, por consiguiente, $\sup\limits_{c \in C} \lambda_0 C_0 = 0$. Entonces la desigualdad (2) adoptará la siguiente forma:

$$\sum_{0}^{m} \lambda_i b_i \geqslant 0 \quad \forall b \in B. \tag{3}$$

C) Los factores λ_i, $i = 0, \ldots, m$ satisfacen las condiciones de no negatividad. Efectivamente, dado que ya hemos dicho que todo vector de componentes no negativas pertenece a B, el vector $(0, \ldots$ $\ldots, 0, 1, \ldots, 0) \in B$, donde la unidad se sitúa en el i-ésimo lugar. Después de introducir ese punto en la expresión (3) obtenemos que $\lambda_i \geqslant 0$.

D) Los factores λ_i, $i = 1, \ldots, m$ satisfacen las condiciones de no rigidez complementaria. Si $f_i(\hat{x}) = 0$, la igualdad $\lambda_i f_i(\hat{x}) = 0$ será trivial. Supongamos que $f_i(\hat{x}) \neq 0$ para cierto i, entonces $f_i(\hat{x}) < 0$. El punto $(0, \ldots, 0, f_i(\hat{x}), 0, \ldots, 0) \in B$, donde el número $f_i(\hat{x})$ se sitúa en el i-ésimo lugar (en (1), en vez de x es suficiente introducir el punto $\hat{x}$). Tras introducir ese punto en la expresión (3) obtenemos $\lambda_i f_i(\hat{x}) \geqslant 0$ y, por consiguiente, $\lambda_i \leqslant 0$. Pero en el punto C ya hemos demostrado que $\lambda_i \geqslant 0$. Por eso $\lambda_i = 0$ y $\lambda_i f_i(\hat{x}) = 0$.

E) En el punto $\hat{x}$ se cumple el principio de mínimo. Tomemos $x \in A$, entonces el punto $(f_0(x), f_1(x), \ldots, f_m(x)) \in B$. Después de introducir ese punto en la expresión (3) obtenemos

$$\sum_{i=0}^{m} \lambda_i f_i(x) = \mathcal{L}(x, \lambda) \geqslant 0.$$

Ahora, si tomamos en consideración la igualdad $f_0(\hat{x}) = 0$ y las condiciones de no rigidez complementaria, entonces, para cualquier $x \in A$ obtendremos

$$\mathcal{L}(x, \lambda) \geqslant 0 = \sum_{i=0}^{m} \lambda_i f_i(\hat{x}) = \mathcal{L}(\hat{x}, \lambda).$$

La afirmación 1 del teorema está demostrada.

F) Demostremos la afirmación 2. Sea $\lambda_0 \neq 0$. Supongamos que $\lambda_0 = 1$. Entonces obtenemos, para todo x admisible,

$$f_0(x) \overset{c)}{\geqslant} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \overset{\text{def}}{=} \mathcal{L}(x, \lambda) \overset{a)}{\geqslant} \mathcal{L}(\hat{x}, \lambda) \overset{\text{def}}{=} f_0(\hat{x}) + \sum_{i=1}^{m} \lambda_i f_i(\hat{x}) \overset{b)}{=} f_0(\hat{x}),$$

es decir, $\hat{x}$ es la solución del problema.

G) Demostremos la afirmación 3. Supongamos que se cumple la condición de Slater, pero en este caso $\lambda_0 = 0$ en la afirmación 1.

Entonces surge inmediatamente la siguiente contradicción: como no todos los factores de Lagrange son iguales a cero, resulta que

$$\mathscr{L}(\bar{x}, \lambda) = \sum_{i=1}^{m} \lambda_i f_i(\bar{x}) < 0 = \mathscr{L}(\hat{x}, \lambda),$$

mientras que, a consecuencia de a), $\mathscr{L}(\bar{x}, \lambda) \geqslant \mathscr{L}(\hat{x}, \lambda)$). $\triangleright$

2.4. Problemas suaves con acotaciones del tipo de igualdades y desigualdades

Examinemos el problema siguiente:

$$f_0(x) \to \inf; \; f_i(x) \leqslant 0, \; i = 1, \ldots, m, \; F(x) = 0, \qquad \text{(Pr)}$$

donde $f_i : X \to \mathbf{R}$, $i = 0, 1, \ldots, m$, $F: X \to Y$, X, Y son espacios normalizados.

Teorema. *Supongamos que X e Y en* (Pr) *son ciertos espacios banachianos (condición de banachianidad); $f_i \in SD(\hat{x})$, $i = 0, 1, \ldots, m$, $F \in SD(\hat{x})$ (condición de suavidad); e* $\operatorname{Im} F'(\hat{x})$, *un subespacio cerrado en Y (condición débil de regularidad). Entonces, si $\hat{x} \in$* locmín Pr, *existirá un vector $\lambda \in \mathbf{R}^{m+1}$ y una funcional $y^* \in Y^*$ que, sin ser iguales simultáneamente a cero, serán tales que para la función de Lagrange del problema* (Pr),

$$\mathscr{L}(x, y^*, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x) + \langle y^*, F(x) \rangle,$$

se cumplirán las condiciones:

a) *de estacionaridad*

$$\mathscr{L}_x(\hat{x}, y^*, \lambda) = 0 \iff \sum_{i=0}^{m} \lambda_i f'_i(\hat{x}) + (F'(\hat{x}))^* y^* = 0;$$

b) *de no rigidez complementaria*

$$\lambda_i f_i(\hat{x}) = 0, \quad i = 1, \ldots, m;$$

c) *de no negatividad*

$$\lambda_i \geqslant 0; \quad i = 0, 1, \ldots, m.$$

Hemos formulado el teorema para el caso de dimensión infinita, con vistas a utilizarlo en esta parte del libro (en el § 5). La demostración se basa en dos afirmaciones referentes al análisis funcional, las cuales se someten a demostración en la segunda parte del libro (p. 7.2). Quienes deseen analizar este teorema en la variante de dimensión finita, deberán tomar en consideración aquellos dos lemas de álgebra superior que han de utilizarse en vez de los lemas del p. 7.2.

Notemos que la formulación del teorema (p. 2.2.2) carece de requisitos acerca del carácter cerrado del subespacio $\operatorname{Im} F'(\hat{x})$. Eso se explica por el hecho de que el subespacio de un espacio de dimensión finita siempre es cerrado.

Pasemos a la demostración del teorema.

◁ Podemos estimar que $f_0(\hat{x}) = 0$, de lo contrario examinemos la función $f_0(x) = f_0(\hat{x}) - f_0(\hat{x})$. Si $f_i(x) \neq 0$ para $1 \leqslant i \leqslant m$, entonces debemos rechazar esa acotación, ya que, para un extremo local, la acotación $f_i(x) < 0$ no tienen importancia. Por esta razón podemos considerar que ya se han cumplido las condiciones de no rigidez complementaria.

A) Caso degenerado. Si $\operatorname{Im} F'(\hat{x})$ es un subespacio propio de Y, entonces, según el lema de no trivialidad del anulador (p. 7.2), existirá una funcional $y^* \in Y^*$, $y^* \neq 0$ tal que $\langle y^*, y\rangle = 0\ \forall y \in \in \operatorname{Im} F'(\hat{x}) \Leftrightarrow \langle y^*, F'(\hat{x})[x]\rangle = 0 \forall x \in X \Leftrightarrow (F'(\hat{x}))^* y^* = 0$. Sólo queda suponer que $\lambda_j = 0$, $j = 0, 1, \ldots, m$ y llegamos a la afirmación del teorema.

B) Supongamos que $F'(\hat{x})$ aplica X a todos los espacios Y, es decir, $\operatorname{Im} F'(\hat{x}) = Y$. Sea, para $0 \leqslant k \leqslant m$,

$$A_k = \{x \mid \langle f_i'(\hat{x}), x\rangle < 0,\ i = k, k+1, \ldots, m,\ F'(\hat{x})[x] = 0\}.$$

Es obvio que $A_0 \subset A_1 \subset \ldots \subset A_m$.

Lema 1 (principal). *Si* $\hat{x} \in$ locmín Pr, *entonces* A_0 *será un conjunto vacío.*

◁◁ Supongamos lo contrario, es decir, que $A_0 \neq \varnothing$. Entonces existirá un vector $\xi \in \operatorname{Ker} F'(\hat{x})$ para el cual $\langle f_i'(\hat{x}), \xi\rangle = \beta_i < 0$, $i = 0, 1, \ldots, m$. Según el teorema de Lusternik* (p. 1.4.4), existe una aplicación $r : [-\alpha, \alpha] \to X$ $(\alpha > 0)$ y un número K tales que

$$F(\hat{x} + \lambda\xi + r(\lambda)) = 0\quad \forall\lambda \in [-\alpha, \alpha],$$

$$\|r(\lambda)\| \leqslant K \|F(\hat{x} + \lambda\xi) - F(\hat{x})\| = K \|\lambda F'(\hat{x})[\xi] + o(\lambda)\| = o(\lambda). \quad (1)$$

Para pequeños $\lambda > 0$ tenemos las siguientes desigualdades (cuando $i = 0, 1, \ldots, m$):

$$f_i(\hat{x} + \lambda\xi + r(\lambda)) = f_i(\hat{x}) + \lambda\langle f_i'(\hat{x}), \xi\rangle + o(\lambda) = \lambda\beta_i + o(\lambda) < 0. \quad (1_i)$$

Las relaciones 1 y 1_i, $i = 1, \ldots, m$ testimonian que para pequeños $\lambda > 0$ el elemento $\hat{x} + \lambda\xi + r(\lambda)$ es admisible en el problema (Pr). Pero en este caso resulta que la desigualdad (1_0) contradice el hecho de que $\hat{x} \in$ locmín Pr.

* En el caso de dimensión finita, según el teorema de función implícita.

C) **Lema 2.** *Si A_m es un conjunto vacío, el principio de Lagrange será válido para el problema* (Pr).

$\triangleleft\triangleleft$ Como $A_m \overset{\text{def}}{=} \{x \mid \langle f'_m(\hat{x}), x\rangle < 0,\ F'(\hat{x})[x] = 0\}$ es un conjunto vacío, por lo tanto, $\langle f'_m(\hat{x}), x\rangle = 0$ para cualquier $x \in \operatorname{Ker} F'(\hat{x})$, es decir, $f'_m(\hat{x})\ (\operatorname{Ker} F'(\hat{x}))^{\perp}$. Dado que $(\operatorname{Ker} F'(\hat{x}))^{\perp} = \operatorname{Im}(F'(\hat{x}))^*$, según el lema acerca del anulador del núcleo de un operador regular (p. 7.2), existirá $y^* \in Y^*$ para el que

$$f'_m(\hat{x}) + (F'(\hat{x}))^* y^* = 0.$$

Hemos obtenido la condición de estacionaridad de la función de Lagrange $\mathcal{L}(x, y^*, \lambda)$ con $\lambda_0 = \ldots = \lambda_{m-1} = 0$, $\lambda_m = 1$. $\triangleright\triangleright$

Así pues, de los puntos B) y C) se deduce que el principio de Lagrange ya está argumentado $(A_m = \varnothing)$ o que

$$\exists k,\ 0 \leqslant k < m:\ A_k = \varnothing,\ A_{k+1} = \varnothing. \tag{2}$$

D) **Lema 3.** *Si se cumplen las relaciones* (2), *entonces $\bar{x} = 0$ será la solución del siguiente problema:*

$$\langle f'_k(\hat{x}), x\rangle \to \inf;\quad \langle f'_i(\hat{x}), x\rangle \leqslant 0,$$
$$i = k+1, \ldots, m,\ F'(\hat{x})[x] = 0. \tag{Pr}$$

$\triangleleft\triangleleft$ Supongamos que la afirmación del lema es incorrecta. Entonces se hallará un elemento η tal que $\langle f'_i(\hat{x}), \eta\rangle \leqslant 0$, $i = k + 1, \ldots, m$, $F'(\hat{x})[\eta] = 0$, pero en este caso $\langle f'_k(\hat{x}), \eta\rangle < 0$. Sea ξ un elemento perteneciente a A_{k+1}, es decir, $\langle f'_i(\hat{x}), \xi\rangle < 0$, $i = k+1, \ldots, m$, $F'(\hat{x})[\xi] = 0$. Entonces, el elemento $\eta + \varepsilon\xi$ con pequeño $\varepsilon > 0$ pertenecerá a A_k en contradicción con (2). $\triangleright\triangleright$

E) Conclusión de la demostración. Apliquemos al problema (Pr) el teorema de Kuhn—Tucker (p. 2.3.3) tomando en consideración que la condición de Slater para ese problema se cumple (debido a la no vacuidad de A_{k+1}). Según este teorema se hallarán ciertos números negativos $\lambda_k = 1, \lambda_{k+1}, \ldots, \lambda_m$ que contribuirán a que para la función de Lagrange del problema (Pr) $\mathcal{L}(x, \lambda) = \sum_{i=k}^{m} \lambda_i < \langle f'_i(\hat{x}), x\rangle$ se cumpla el principio de mínimo en el punto $\bar{x} = 0$:

$$\min_{x \in \operatorname{Ker} F'(\hat{x})} \widetilde{\mathcal{L}}(x, \lambda) = \widetilde{\mathcal{L}}(\bar{x}, \lambda) = 0.$$

De esta última relación se desprende que

$$\widetilde{\mathcal{L}}(x, \lambda) = \Big\langle \sum_{i=k}^{m} \lambda_i f'_i(\hat{x}), x\Big\rangle = 0 \text{ para cualquier}$$
$$x \in \operatorname{Ker} F'(\hat{x}),$$

es decir,

$$\sum_{i=k}^{m} \lambda_i f_i'(\hat{x}) \in (\operatorname{Ker} F'(\hat{x}))^{\perp}.$$

En vista de que $(\operatorname{Ker} F'(\hat{x}))^{\perp} = \operatorname{Im}(F'(\hat{x}))^*$, según el lema del anulador del núcleo del operador regular, existirá $y^* \in Y^*$ para el cual

$$\sum_{i=k}^{m} \lambda_i f_i'(\hat{x}) + (F'(\hat{x}))^* y^* = 0.$$

Pero esta expresión es precisamente la condición de estacionaridad de la función de Lagrange $\mathcal{L} \times (x, y^*, \lambda)$, suponiendo que $\lambda_0 = \ldots = \lambda_{k-1} = 0$.

Nota. De la demostración del teorema se deduce que si $\operatorname{Im} F'(\hat{x}) = Y$ (es decir, si F es regular en $\hat{x}$) y existe un elemento $h \in \operatorname{Ker} F'(\hat{x})$ para el cual $\langle f_i'(\hat{x}), h\rangle < 0$, $i = 1, \ldots, m$ (llamemos esta condición análogo de la condición de Slater), entonces $\lambda_0 \neq 0$ y, por consiguiente, podemos suponer que $\lambda_0 = 1$.

2.5. Ejemplos

1. $f(x) = f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 - 2x_1 + x_2 \to \text{extr}$.

Solución. Condición necesaria de extremo de primer orden de una función suave:

$$f'(\hat{x}) = 0 \Leftrightarrow \frac{\partial f(\hat{x})}{\partial x_1} = \frac{\partial f(\hat{x})}{\partial x_2} = 0 \Leftrightarrow \begin{cases} 2x_1 - x_2 - 2 = 0; \\ -x_1 + 2x_2 + 1 = 0. \end{cases}$$

Resolviendo este sistema de ecuaciones hallamos el único punto estacionario $\hat{x} = (\hat{x}_1, \hat{x}_2) = (1, 0)$. Para verificar las condiciones de segundo orden escribimos la matriz de segundas derivadas:

$$\left(\frac{\partial^2 f(\hat{x})}{\partial x_i\, \partial x_j}\right)_{i,\, j=1,\, 2} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

De acuerdo con el criterio de Sylvester (p. 2.1.2), esa matriz es definida positivamente. En virtud de la condición de suficiencia del extremo de la función de varias variables (p. 2.1.2), $(1, 0) \in \text{locmín}\, f$. Es fácil comprender que en realidad $(1, 0) \in \text{absmín}\, f$, mientras que $S_{\text{máx}} = +\infty$.

2. $4x_1 + 3x_2 \to \text{extr}$; $x_1^2 + x_2^2 = 1$.

Solución. Recurrimos a la regla de factores lagrangianos para resolver los problemas suaves con acotaciones del tipo de igualdades (P. 2.2.2). La función de Lagrange:

$$\mathcal{L} = \lambda_0 (4x_1 + 3x_2) + \lambda (x_1^2 + x_2^2 - 1).$$

La condición necesaria:

$$\mathcal{L}_x = 0 \Leftrightarrow 4\lambda_0 + 2\lambda x_1 = 0, \quad 3\lambda_0 + 2\lambda x_2 = 0.$$

Si $\lambda_0 = 0$, entonces $\lambda \neq 0$. Por consiguiente, partiendo de las ecuaciones anteriores resulta que $x_1 = x_2 = 0$. El punto $(0, 0)$ no es admisible. Supongamos que $\lambda_0 = 1$. Entonces $x_1 = -2/\lambda$, $x_2 = -3/2\lambda$. Introduciendo x_1 y x_2 en la acotación $x_1^2 + x_2^2 = 1$, resulta que $\lambda = \pm 5/2$ y obtenemos, respectivamente, dos puntos estacionarios $(4/5, 3/5)$, $(-4/5, -3/5)$.

Según el teorema de Weierstrass existen soluciones de los problemas de máximo y mínimo. Al considerar los valores de la funcional en los puntos estacionarios, obtenemos

$$(4/5,\ 3/5) \in \text{absmáx},\ S_{\text{máx}} = 5,$$

$$(-4/5,\ -3/5) \in \text{absmín},\ S_{\text{mín}} = -5.$$

3. $x_1^2 + x_2^2 + x_3^2 \to \inf;\quad 2x_1 - x_2 + x_3 \leqslant 5,\quad x_1 + x_2 + x_3 = 3.$

Solución. Apliquemos la regla de solución de problemas suaves con acotaciones del tipo de igualdades y desigualdades (p. 2.4). La función de Lagrange: $\mathscr{L} = \lambda_0 (x_1^2 + x_2^2 + x_3^2) + \lambda_1 (2x_1 - x_2 + x_3 - 5) + \lambda_2 (x_1 + x_2 + x_3) - 3)$.

Las condiciones necesarias:

a) de estacionaridad

$$\mathscr{L}_{x_1} = 0 \Leftrightarrow 2\lambda_0 x_1 + 2\lambda_1 + \lambda_2 = 0,$$

$$\mathscr{L}_{x_2} = 0 \Leftrightarrow 2\lambda_0 x_2 + \lambda_2 - \lambda_1 = 0,$$

$$\mathscr{L}_{x_3} = 0 \Leftrightarrow 2\lambda_0 x_3 + \lambda_2 + \lambda_1 = 0;$$

b) de no rigidez complementaria

$$\lambda_1 (2x_1 - x_2 + x_3 - 5) = 0;$$

c) de no negatividad $\lambda_0 \geqslant 0$, $\lambda_1 \geqslant 0$.

$\lambda_0 = 0 \overset{a)}{\Longrightarrow} \lambda_1 = \lambda_2 = 0$ todos los factores de Lagrange son ceros. Sea $\lambda_0 = 1/2$. Supongamos que $\lambda_1 \neq 0 \overset{b)}{\Longrightarrow} 2x_1 - x_2 + x_3 - 5 = 0$. Expresemos los valores de x_1, x_2 y x_3 de la condición a) mediante λ_1 y λ_2 e, introduciéndolos en las ecuaciones $x_1 + x_2 + x_3 = 3$, $2x_1 - x_2 + x_3 - 5 = 0$, obtenemos $\lambda_1 = -9/14 < 0$, lo cual contradice la condición c). Entonces $x_1 = x_2 = x_3 = 1$ es un punto crítico.

La función $f(x) = x_1^2 + x_2^2 + x_3^2 \to +\infty$ para $|x| \to \infty$, por consiguiente, según el corolario del teorema de Weierstrass (p. 1.2.1), la solución del problema es real y, en virtud de la unicidad del punto crítico, esa solución sólo puede ser una. Así, $\hat{x} = (1, 1, 1) \in \text{absmín}$, $S_{\text{mín}} = 3$.

4. $\langle Ax, x\rangle \to \inf;\ \langle x, x\rangle = 1$ ($x \in \mathbf{R}^n$, $A = (a_{ij})_{i,j=1}^n$ es una matriz simétrica).

Solución. La existencia de la solución de $\hat{x}$ resulta del teorema de Weierstrass, ya que la esfera $S^{n-1} = \{x \in \mathbf{R}^n \mid \times \mid x|^2 \overset{\text{def}}{=} \langle x, x\rangle = 1\}$ es compacta. La función de Lagrange:

$$\mathscr{L} = \lambda_0 \langle Ax, x\rangle + \lambda_1 \langle x, x\rangle.$$

Las condiciones necesarias:

$$\mathscr{L}_x (x, \lambda_0, \lambda_1) = 0 \Leftrightarrow \lambda_0 A_x + \lambda_1 x = 0, \quad \lambda_0 \geqslant 0.$$

Si $\lambda_0 = 0$, entonces $\lambda_1 \neq 0$ y, por lo tanto, $\hat{x} = 0$, lo cual contradice la ecuación de relación $\langle x, x\rangle = 1$. Sea $\lambda_0 = 1$. En este caso $A\hat{x} = -\lambda_1 \hat{x}$. Por consiguiente, la solución será el propio vector de la matriz A. Después de multiplicar complementariamente $A\hat{x} = -\lambda_1 \hat{x}$ por $\hat{x}$ obtenemos $S_{\text{pr}} = -\lambda_1$. Con otras palabras, la solución del problema de mínimo es el propio vector de la matriz A, correspondiente a su propio valor mínimo.

2.6. Acerca de los métodos de solución de problemas extremales. Método de gradiente y método de Newton

Aquí nos limitaremos al caso elemental de minimización incondicional de dimensión finita, es decir, a la minimización de la función suave f en un problema no acotado:

$$f(x) \to \inf; \quad x \in \mathbf{R}^n. \qquad \text{(Pr)}$$

Anteriormente hemos prestado mucha atención a las condiciones de extremo necesarias.

La teoría de problemas extremales propone la siguiente receta para resolver el problema Pr. Es preciso hallar todos los puntos que satisfagan la condición necesaria de primer orden (es decir, el teorema de Fermat)

$$f'(x) = 0, \qquad (1)$$

y después verificarlos con arreglo a su máximo y mínimo recurriendo a los criterios de segundo orden. Pero cabe señalar que, al resolver problemas prácticos concretos utilizando ordenadores, los métodos de búsqueda de extremos no siempre se hallan directamente enlazados con la teoría de problemas extremales. Dicha teoría permite comprender mejor los planteamientos de tales problemas, las posibilidades que ellos encierran, etc. Mostremos que también existen otros métodos muy efectivos que difieren (al aplicarlos, por ejemplo, al problema de minimización incondicional) de la solución de la ecuación (1).

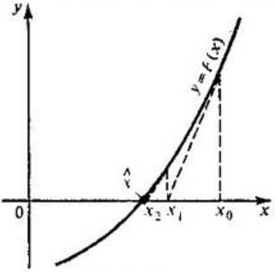


Fig 2

Pero comencemos, no obstante, por el método antiguo, perteneciente a Newton y destinado precisamente a la solución de la ecuación (1).

Sea $F\colon \mathbf{R}^n \to \mathbf{R}^n$ una aplicación de clase C^1. Para resolver la ecuación $F(x) = 0$ es aplicable el siguiente procedimiento. Se toma cierto punto x^0 y luego se construye la sucesión de puntos según la regla

$$x^{k+1} = x^k - (F'(x^k))^{-1} F(x^k), \quad k = 0, 1, \ldots, \qquad (2)$$

Este procedimiento para el caso de $n = 1$ se expone en la fig. 2.

Para hallar los puntos de mínimo de la función $f \in C^2(\mathbf{R}^n)$ conforme a la expresión (1), es necesario usar el procedimiento (2) siempre que se elija x^0 en calidad de aproximación inicial. Entonces llegamos al siguiente proceso iterativo, denominado *método de Newton*:

$$x^{k+1} = x^k - (f''(x^k)^{-1} f'(x^k), \quad k = 0, 1, \ldots \qquad (3)$$

Se puede mostrar que si f'' satisface la condición de Lipschitz y la de convexidad fuerte ($f''(x) \geqslant \alpha 1$) y la norma $f'(x^0)$ es bastante pequeña, el método 3 convergirá al punto $\hat{x}$ de mínimo global f con una velocidad cuadrática de ($\| x^k - \hat{x} \| \leqslant Cq^{2^k}$).

El método sujeto a examen no está relacionado directamente con la solución de la ecuación (1). El mismo consiste en lo siguiente. Supongamos que $f \in C^1(\mathbf{R}^n)$. Entonces será necesario elegir de nuevo cierto punto inicial x^0 y construir la sucesión de puntos conforme a la regla

$$x^{k+1} = x^k - \alpha^k f'(x^k), \quad \alpha^k > 0, \quad k = 0, 1, \ldots \tag{4}$$

Tal tipo de procedimientos se denominan *métodos gradientales*. El número α^k se llama *longitud del paso del método*.

La idea del método es sencilla. El vector $f'(x)$ tiende al máximo crecimiento de la función f en el punto x, mientras que el vector $-f$ se halla orientado hacia el lado de máximo decrecimiento. Cada vez que realizamos la iteración (4) «descendemos» por la funcional. Por eso se puede esperar que nos aproximemos al punto de mínimo.

La variante elemental del método gradiental surge cuando $\alpha^k \equiv \equiv \alpha = \text{const}$ en (4). Se puede mostrar que si el gradiente f' satisface la condición de Lipschitz, f se halla acotada por abajo y α es bastante pequeño; además, en este caso $f'(x^k) \to 0$ cuando $k \to \infty$. A su vez, si suponemos que f es una función muy convexa, se pondrá de manifiesto el hecho de que x^k tiende al punto de mínimo global f a velocidad de progresión geométrica.

La comparación de los dos métodos descritos refleja sus rasgos fuertes y débiles. Los rasgos fuertes del método gradiental son: la sencillez de los cálculos y la pequeña suavidad aplicada a f. Pero ese método converge más lentamente y requiere experimentos para elegir α. Sin embargo, el método de Newton converge con gran rapidez, pero son bastante estrictas las exigencias de suavidad de la función. Además, el mismo requiere un número de cálculos mucho mayor. Todo esto se examina más detalladamente en [5, 16 y 21], así como en la monografía de F. P. Vasíliev «Métodos de solución de problemas extremales» (Moscú, Ed. Naúka, 1981).

§ 3. Problemas de programación lineal

La programación lineal es una de las partes mejor elaboradas teóricamente de los métodos de optimización. En ella se estudian los problemas acerca del extremo de la función lineal con acotaciones del tipo de igualdades y desigualdades, también dadas en forma de funciones lineales. Este tipo de problemas desempeñan un papel muy importante en las aplicaciones prácticas, sobre todo en la economía. En el presente párrafo se examinarán los aspectos fundamentales de la referida teoría: las cuestiones relacionadas con la existencia de soluciones, los criterios de la solución, los algoritmos numéricos y la teoría de dualidad.

3.1. Método simplex

3.1.1. Planteamiento del problema. Llamaremos problema general de programación lineal el problema

$$\langle c, x\rangle \to \inf: \ Ax \leqslant b. \tag{Pr}$$

Los problemas de programación lineal se analizan, por regla general, reduciéndolos a la forma canónica

$$\langle c, x\rangle \to \sup: \ Ax = b, \ x \geqslant 0, \tag{$\mathrm{Pr_c}$}$$

o a la forma normal

$$\langle c, x\rangle \to \sup: \ Ax \leqslant b, \ x \geqslant 0. \tag{$\mathrm{Pr_n}$}$$

Aquí $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, $c = (c_1, \ldots, c_n) \in \mathbf{R}^n$, $b = (b_1, \ldots, b_m) \in$ $\in \mathbf{R}^m$, $A = (a_i^j)_{\substack{i=1, \ldots, m \\ j=1, \ldots, n}} = \begin{pmatrix} a_1^1 & \ldots & a_1^n \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ a_m^1 & \ldots & a_m^n \end{pmatrix}$ es una matriz de dimensiones $m \times n$ con columnas $a^j = (a_1^j, \ldots, a_m^j)$ $j = 1, \ldots, n$.

La forma canónica es más cómoda para describir los algoritmos de solución, además, los problemas (Pr) y ($\mathrm{Pr_n}$) se utilizan a menudo al examinar las cuestiones de existencia de soluciones y de dualidad. El problema de forma normal siempre puede ser reducido a una forma canónica o general introduciendo coordenadas complementarias y cambiando la matriz A. Pero también son posibles resultados inversos. Por ejemplo, si se ha dado el problema $\langle c, x\rangle \to \sup$; $Ax \leqslant b$, $(x \geqslant 0$, es posible introducir las coordenadas $\bar{x} = (x_{n+1}, \ldots, x_{n+m})$ y obtener un problema de forma canónica:

$$\langle c, x\rangle \to \sup; \ \langle a^j, x\rangle + x_{n+j} = b_j,$$
$$x \geqslant 0, \ \bar{x} \geqslant 0.$$

Designemos por C el conjunto de puntos admisibles en un problema de programación lineal. Ese conjunto es poliedro convexo en el espacio $\mathbf{R}^n$. Es fácil notar que el extremo de la función lineal (si éste existe) puede ser alcanzado en el punto extremo (angular) del poliedro convexo. El número de puntos extremos del conjunto C, que se dan en forma de un número finito de igualdades y desigualdades lineales, es un número finito [20, pág. 190]. Por lo tanto, para resolver un problema de programación lineal (si esa solución existe), es suficiente examinar los valores de la función $\langle c, x\rangle$ en todos los puntos extremos del conjunto C. Pero la búsqueda de todos esos puntos y el estudio de los valores de la función es una operación bastante compleja. El método de solución del problema de programación lineal (método símplex) que se expone más abajo permite, partiendo de cierto punto extremo inicial, pasar a otro punto en dirección del máximo crecimiento del valor de la función.

El problema (Pr_c) se llama no degenerado si todo punto extremo del conjunto C contiene exactamente m coordenadas positivas.

Sea x el punto extremo en un problema no degenerado con m coordenadas positivas $x_1, \ldots, x_m$ (supongamos que son las primeras). Entonces el vector x puede ser representado en forma de $x = (x_{\text{bás}}, x_{\text{no bás}})$, donde $x_{\text{bás}} = (x_1, \ldots, x_m)$ es el vector básico; y $x_{\text{no bás}} = (0, \ldots, 0) \in \mathbf{R}^{n-m}$, el vector no básico. Análogamente la matriz A puede ser representada en forma de $A = (A_{\text{bás}}, A_{\text{no bás}})$ y así demostremos (p. 3.2.3) que la matriz $A_{\text{bás}}$ no es degenerada.

3.1.2. Regla de solución. Para resolver un problema no degenerado de programación lineal es necesario:

1. Reducir el problema a la forma canónica (Pr_c).

2. Localizar el punto extremo $x_0 = (x_1, \ldots, x_m, 0, \ldots, 0) \in \mathbf{R}^n$, $x_i > 0$, $i = 1, \ldots, m$ en el conjunto de elementos admisibles C. El método de búsqueda del punto extremo inicial será descrito más abajo.

3. Hacer la tabla simplex para el punto extremo inicial x_0. Aclaraciones respecto a esa tabla: la misma contiene $n + 4$ columnas y $m + 4$ renglones.

En la primera columna, desde el tercero hasta el m + segundo lugar se encuentran los vectores básicos $a_1, \ldots, a_m$ correspondientes al punto extremo inicial $x_0 = (x_1, \ldots, x_m, 0, \ldots, 0)$ sujeto a examen.

En la segunda columna, en lugares análogos están situados los valores c_j del vector c con los mismos números que las columnas a^j.

En el primer renglón, partiendo de la cuarta columna, se encuentran los elementos $c_1, \ldots, c_n$.

En el segundo renglón, partiendo de la tercera columna, se encuentran los vectores $b, a^1, \ldots, a^n$. Debajo de ellos se ofrece su desarrollo con arreglo a la base $a^1, \ldots, a^m$. Está claro que $b = \sum_{i=1}^{m} a^i x_i$, $a^i = a^i$, $i = 1, \ldots, m$ y $a^j = \sum_{i=1}^{m} a^i x_{ij}$, $j = m + 1, \ldots, n \Leftrightarrow a^j = A_{\text{bás}} x^j \Leftrightarrow x^j = A_{\text{bás}}^{-1} a^j$.

Escribamos $z_0 = \langle c_{\text{bás}}, x_0 \rangle$ en el penúltimo renglón de la columna debajo del vector b; y $z_i = \langle c_{\text{bás}}, x^i \rangle$, debajo de los vectores c_i, $i = 1, \ldots, n$. Está claro que z_0 es el valor de la funcional en el punto extremo inicial x_0, $z_i = c_i$, $i = 1, \ldots, m$. En el último renglón se escribe la diferencia entre los elementos del penúltimo renglón y los elementos del primer renglón: $\Delta_i = z_i - c_i$, $i = 1, \ldots, n$.

4. Analizar la tabla simplex.

a) Si $\Delta \geqslant 0$, el punto extremo x_0 será la solución del problema.

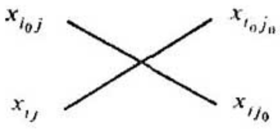
b) Si $\Delta_j < 0$ y $x^j \leqslant 0$ para cierto j, el valor del problema será $S_{\text{máx}} = +\infty$.

c) Supongamos que el renglón Δ tiene números negativos y que las columnas correspondientes x^j contienen números positivos.

Supongamos que mín $\Delta_j = \Delta_{j_0} < 0$. Evidentemente que $m + 1 \leqslant j_0 \leqslant n$. La columna correspondiente al índice j_0 se llama *columna de resolución*. Si sín Δ_j es alcanzado por varios valores de j, en calidad de columna de resolución elegiremos la columna con cualquier índice de ese tipo. Denotemos $\theta_i = x_i/x_{ij_0}$ para $x_{ij_0} > 0$, $i = 1, \ldots, m$, $\theta_0 =$ mín $\{\theta_i/x^{ij_0} > 0, i = 1, \ldots, m\} = \theta_{i_0} > 0$. El renglón del vector a^{i_0} se denomina *renglón de resolución*. Si mín θ_i es alcanzado por varios valores de i, en calidad de renglón de resolución elegiremos el renglón con cualquier índice de ese tipo. El elemento $x_{i_0j_0}$ situado en la columna de resolución y en el renglón de resolución se llama *elemento de resolución* de la tabla.

Seguidamente se ha de eliminar el vector a^{i_0} del número de vectores básicos, y en su lugar tomar el vector a^{j_0}. El valor de la funcional en el nuevo punto extremo x_{nuevo} con nuevos números básicos $1, \ldots, i_0 - 1, j_0, i_0 + 1, \ldots, m$ crecerá en $-\theta_0\Delta_{j_0}$.

5. Hacer una nueva tabla simplex para la nueva base $a^1, \ldots, a^{i_0-1}, a^{j_0}, a^{i_0+1}, \ldots, a^{i_0}$, es decir, de hecho, desarrollar los vectores $b, a^1, \ldots, a^n$ con arreglo a esa nueva base. Señalemos, sin argumentación (lo cual será realizado en el p. 3.2.3), la manera de hacer la nueva tabla simplex según la tabla anterior. Los elementos x_{ij} de dicha tabla, situados debajo de los vectores $b, a^1, \ldots, a^n$ y que no se hallan en la columna de resolución o en el renglón de resolución de la tabla simplex anterior, se calculan por la regla del rectángulo:



del número x_{ij} se resta el producto de x_{i_0j} por x_{ij_0}, dividido entre $x_{i_0j_0}$. Los elementos del renglón de resolución se calculan por la fórmula $(x_{i_0j_0})_{\text{nuevo}} = x_{i_0j}/x_{i_0j_0}$, $j = 1, \ldots, m$. Está claro que los demás elementos de la columna de resolución $(x_{i_0j_0})_{\text{nuevo}} = 1$ son iguales a cero.

Luego debemos analizar de nuevo la tabla simplex, es decir, regresar al p. 4, etc., hasta que lleguemos a la solución del problema.

3.1.3. Ejemplo. Resolver el problema no degenerado de programación lineal en la forma canónica

$$2x_1 + x_2 + x_3 - x_4 \to \sup; \ x_1 - x_2 + x_3 = 1,$$
$$x_i \geqslant 0, \ i = 1, 2, 3, 4, \ 2x_1 + x_2 + x \ = 3$$

con un punto extremo inicial dado $x_0 = (0, 0, 1, 3)$.

	c		c_1	$\cdots$	c_m	c_{m+1}	$\cdots$	c_{j_0}	$\cdots$	c_n	
base		b	a^1	$\cdots$	a^m	a^{m+1}	$\cdots$	a^{j_0}	$\cdots$	a^n	θ
a^1	c_1	x_1	1	$\cdots$	0	x_{1m+1}	$\cdots$	x_{1j_0}	$\cdots$	x_{1n}	θ_1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
a^{i_0}	c_{i_0}	x_{i_0}	0	$\cdots$	0	x_{i_0m+1}	$\cdots$	$x_{i_0j_0}$	$\cdots$	x_{i_0n}	θ_{i_0}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	.	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
a^m	c_m	x_m	0	$\cdots$	1	x_{mm+1}	$\cdots$	x_{mj_0}	$\cdots$	x_{mn}	θ_m
z		$\langle c_{\text{bás}} x_0\rangle$	c_1	$\cdots$	c_m	$\langle c_{\text{bás}} x^{m+1}\rangle$	$\cdots$	$\langle c_{\text{bás}} x_{j_0}\rangle$	$\cdots$	$\langle c_{\text{bás}} x^n\rangle$	
$\Delta = z - c$			0	$\cdot\cdot$	0	$\langle c_{\text{bás}} x^{m+1}\rangle - c_{m+1}$	$\cdots$	$\langle c_{\text{bás}} x^{j_0}\rangle - c_{j_0}$	$\cdots$	$\langle c_{\text{bás}},\ x^n\rangle - c_n$	

Solución. Los vectores básicos $a^3 = (1, 0)$ y $a^4 = (0, 1)$. Hagamos la primera tabla simplex:

	c		2	1	1	-1	θ
base		b	a^1	a^2	a^3	a^4	
a^3	1	1	1	-1	1	0	
a^4	-1	3	2	1	0	1	$3\rightarrow$
z		-2	-1	-2	1	-1	
$z-c$			-3	-3	0	0	

De esta tabla se deduce que en calidad de columna de resolución se pueden tomar las columnas a^1 y a^2. Para mayor precisión tomamos la columna a^2. Entonces $\theta = 3$ y el renglón de resolución será a^4. Sustituimos, en la base, el vector a^4 por a^2, y para la nueva base hacemos la segunda tabla simplex:

	c		2	1	1	-1	θ
base		b	a^1	a^2	a^3	a^4	
a^3	1	4	3	0	1	1	
a^2	1	3	2	1	0	1	
z		7	5	1	1	2	
$z-c$			3	0	0	3	

El vector $z - c \geqslant 0$, por eso el punto $\hat{x} = (0, 3, 4, 0)$ será la solución del problema y $S_{\max} = 7$.

Si en calidad de columna de resolución en la primera tabla simplex hubiéramos tomado la columna a^1, llegaríamos a un mismo punto $(0, 3, 4, 0)$, pero deberíamos hacer mayor cantidad de pasos.

3.2. Argumentación del método simplex

3.2.1. Teorema de existencia. Examinemos el problema de programación lineal en su forma general

$$\langle c, x\rangle \to \inf; \quad Ax \leqslant b, \tag{Pr}$$

donde $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, $A = \begin{pmatrix} a_1^1, \dots, a_1^n \\ \cdots\cdots \\ a_m^1, \dots, a_m^n \end{pmatrix}$, S_{Pr} es el valor numérico del problema Pr; y ArgPr, el conjunto de soluciones de ese mismo problema, o sea, el conjunto de puntos admisibles $x \in \mathbf{R}^n$ para los cuales $\langle c, x\rangle = S_{\mathrm{Pr}}$.

Teorema. *Si el valor numérico del problema* Pr *es finito* ($|S_{\mathrm{Pr}}| < \infty$), *existirá su solución* ($\mathrm{ArgPr} \neq \varnothing$).

$\lhd$ Señalemos inmediatamente que, en vista de que el valor numérico del problema es finito, el conjunto de elementos admisibles no es vacuo ($D_{\mathrm{Pr}} \neq \varnothing$).

Analicemos el conjunto

$$K = \{(\alpha, z) \in \mathbf{R} \times \mathbf{R}^m \mid \exists x, \ \langle c, x\rangle \leqslant \alpha, \ Ax \leqslant z\}.$$

Evidentemente que K es un cono convexo.

Lema 1. *K es un cono finitamente generado.*

$\lhd \lhd$ Mostremos que $K = \mathrm{cone}\,\{\pm\xi_1, \dots, \pm\xi_n, \xi_{n+1}, \dots, \xi_{n+m+1}\}$, donde $\xi_i = (c_i, a_1^i, \dots, a_m^i)$, $i = 1, \dots, n$, $\xi_{n+1} = (1, 0, \dots, 0)$, $\xi_{n+2} = (0, 1, \dots, 0)$, $\dots$, $\xi_{n+m+1} = (0, \dots, 0, 1)$.

El encaje $\mathrm{cone}\,\{\pm\xi_1, \dots, \pm\xi_n, \xi_{n+1}, \dots, \xi_{n+m+1}\} \subset K$ se deduce del hecho de que todas las generatrices del cono se encuentran en K. Efectivamente, suponiendo $x = \pm e_i$ en la definición del cono K, obtenemos que $\pm\xi_i \in K$, $i = 1, \dots, n$. Para los demás i hay que tomar $x = 0$.

En cambio, si el vector $\xi = (\alpha, z) = (\alpha, z_1, \dots, z_m) \in K$, entonces para cierto $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ $\langle c, x\rangle \leqslant \alpha$, $Ax \leqslant z$. Eso quiere decir que para ciertos $\beta_0 \geqslant 0, \dots, \beta_m \geqslant 0$ se cumplen las relaciones

$$\sum_{j=1}^n c_j x_j + \beta_0 = \alpha, \quad \sum_{j=1}^n a_i^j x_j + \beta_i = z_i,$$
$$i = 1, \dots, m.$$

Lo cual precisamente significa que

$$\sum_{j=1}^n x_j \xi_j + \beta_0 \xi_{n+1} + \sum_{i=1}^m \beta_i \xi_{n+i+1} = (\alpha, z) = \xi.$$

Como $\sum_{j=1}^n x_j \xi_j = \sum_{j=1}^n |x_j| (\xi_j \operatorname{sign} x_j)$, resulta que

$$\xi \in \mathrm{cone}\,\{\pm\xi_1, \dots, \pm\xi_n, \xi_{n+1}, \dots, \xi_{n+m+1}\}. \rhd\rhd$$

Lema 2 (acerca del carácter cerrado del cono finitamente generado). *El cono finitamente generado es un cono cerrado.*

$\lhd\lhd$ Sea $K = \text{cone}\,\{x_1, \ldots, x_N\} = \{x - \sum_{i=1}^{N} \lambda_i x_i \mid \lambda_i \geqslant 0, i = 1, \ldots, N\}, x_1, \ldots, x_N \in \mathbf{R}^n$.

$\lhd\lhd$ Supongamos que $A\lambda = \sum_{i=1}^{N} \lambda_i x_i$ para el vector $\lambda = (\lambda_1, \ldots \ldots, \lambda_N) \in \mathbf{R}^n$, entonces $A \in \mathcal{L}(\mathbf{R}^N, \mathbf{R}^n)$. La imagen $A\mathbf{R}^n$ (denotémosla por L) es un subespacio de dimensión finita en $\mathbf{R}^n$. Según el lema acerca del operador derecho inverso (p. 7.2), existe un operador inverso acotado $M: L \to \mathbf{R}^N$ tal que

$$A \circ Mx = x \forall x \in L, \quad | Mx | \leqslant C | x |. \tag{1}$$

Supongamos que $\hat{x} \neq 0$ pertenece a la clausura de K ($\hat{x} = 0 \in K$ según la definición), es decir, existe una sucesión $\{\lambda^h\}_{h \in \mathbf{N}}, \lambda^h = (\lambda_1^h, \ldots, \lambda_N^h)$ tal que $A\lambda^h =: x^h \to \hat{x}$. Entonces se hallará $k_0 > 0$ tal que, para $K \geqslant k \mid x^h \mid \leqslant 2\hat{x}$ y de la expresión 1, obtenemos

$$| \lambda^h | \overset{\text{def}}{=} | A^{-1} x^h | \leqslant C | x^h | \leqslant 2C | \hat{x} |.$$

A consecuencia del carácter acotado de $\{\lambda^h\}$, de ésta puede ser elegida la sucesión contenida convergente $\lambda_k^n \to \hat{\lambda}$. Entonces

$$A\bar{\lambda} = \lim_{h \to \infty} A\lambda^{n_k} \overset{\text{def}}{=} \lim_{x \to \infty} x^{n_k} = \hat{x} \overset{\text{def}}{\Rightarrow} \hat{x} \in 1.$$

Está claro que $\hat{x} \in K$, o sea, K es un conjunto cerrado. $\rhd\rhd$

Según la condición, existe el elemento admisible x, es decir, el punto $(\langle c, x \rangle, b) \in K$ Además, el valor numérico del problema $\hat{\alpha} > -\infty$ (en particular, $\langle c, x \rangle \geqslant \alpha$). Todo eso significa que el conjunto $A = \{\alpha \in \mathbf{R}\, (\alpha, b) \in K\}$ no es vacuo y que $\hat{\alpha} = \inf \{\alpha \mid \alpha \in \in A\}$. De aquí se desprende que $(\hat{\alpha}, b)$ pertenece a la clausura de K y, según el lema 2, también a K. Lo cual precisamente significa que existe un valor de $\hat{x}$ para el cual $\langle c, \hat{x} \rangle \leqslant \alpha$, $A\hat{x} \leqslant b$, es decir, $\hat{x}$ es la solución del problema. $\rhd$

Análogamente se demuestran los teoremas de existencia para los problemas de programación lineal planteados de otras formas:

$$| S_{\Pr} | < \infty \Rightarrow \operatorname{Arg} \Pr \neq \varnothing.$$

De ese mismo modo se demuestra que el mínimo existe en la función poliedral $\max\limits_{1 \leqslant i \leqslant N} (\langle c_i, x \rangle + \lambda_i)$ de la variedad afín $\langle b_j, x \rangle = 0$, $1 \leqslant j \leqslant m < n$ $(x \in \mathbf{R}^n)$.

3.2.2. Teoría de dualidad y criterio de solución. Analicemos el problema de programación lineal (p. 3.1.1):

$$\langle c, x\rangle \to \inf; \quad Ax \leqslant b. \tag{Pr}$$

Llámase *dual* a éste, el problema

$$\langle b, y\rangle \to \sup; \quad A^*y = c, \quad y \leqslant 0. \tag{Pr*}$$

La deducción de un problema dual se ofrece en el p. 11.1.1. Si presentamos el problema (Pr*) en forma de (Pr) y le buscamos su problema dual, resultará que el problema dual a (Pr*) será el problema (Pr). Por lo tanto, hemos obtenido un par de problemas duales.

Es fácil deducir que a un problema de forma canónica le será dual el siguiente problema:

$\langle b, y\rangle \to \inf; \quad A^*y \geqslant c.$

Para los problemas de forma normal, los más simétricos resultan los problemas directos y duales:

$$\langle c, x\rangle \to \sup; \; Ax \leqslant b, \; x \geqslant 0,$$
$$\langle b, y\rangle \to \inf; \quad A^* y \geqslant c, \quad y \geqslant 0.$$

Aquí los vectores b y c se cambian de lugares, inf se sustituye por sup y, viceversa, la matriz A se sustituye por una matriz transpuesta y, a su vez, las desigualdades matriciales se reemplazan por desigualdades orientadas en sentidos opuestos.

Teorema 1 (de dualidad). *Para un par de problemas duales de programación lineal tiene lugar la siguiente alternativa: el valor de uno de los problemas es finito (y entonces también se finito el valor del segundo problema y ambos valores coinciden), o bien el conjunto de elementos admisibles es vacuo en uno de los problemas (y entonces el otro problema es incompatible o tiene valor infinito).*

La demostración se ofrece en el p. 11.1.1.

Teorema 2 (criterio de solución). *Sean $\hat{x}$ e $\hat{y}$ los elementos admisibles en los problemas directo* (Pr) *y dual* (Pr*) *respectivamente* ($\hat{x} \in D_{\mathrm{Pr}}$, $\hat{y} \in D_{\mathrm{Pr}^*}$. *Entonces $\hat{x}$ e $\hat{y}$ serán las soluciones en sus problemas respectivos* ($\check{x} \in \operatorname{Arg} \mathrm{Pr}$, $y \in \operatorname{Arg} \mathrm{Pr}^*$) *si y sólo si se cumple la igualdad* $\langle c, \hat{x}\rangle = \langle \hat{y}, b\rangle$.

Necesidad. Sean $\hat{x} \in \operatorname{Arg} \mathrm{Pr}$, $\hat{y} \in \operatorname{Arg} \mathrm{Pr}^*$, lo cual significa que ambos problemas son compatibles y sus valores son finitos. De acuerdo con el teorema de dualidad, esos dos valores son iguales, es decir, $\langle c, \hat{x}\rangle = S_{\mathrm{Pr}} = S_{\mathrm{Pr}^*} = \langle \hat{y}, b\rangle$.

Suficiencia. Sean x, y los elementos admisibles aleatorios en los problemas (Pr) y (Pr*), o sea, $Ax \leqslant b$, $A^*y = c$, $y \leqslant 0$.

Entonces

$$\langle c, x\rangle = \langle A^*y, x\rangle = \langle y, Ax\rangle \geqslant \langle y, b\rangle. \tag{1}$$

De la desigualdad (1) se deduce que $\langle c, x\rangle \geqslant \langle \check{y}, b\rangle$ para cualquier $x \in D_{\mathrm{Pr}}$. Nosotros tenemos $\langle c, \hat{x}\rangle = \langle \hat{y}, b\rangle$, por lo tanto, $\hat{x}$

es la solución del problema (Pr). Análogamente se deduce que $\hat{y}$ es la solución del problema (Pr*). ▷

No es difícil comprender que el teorema 2 es válido no sólo para los problemas (Pr) y (Pr*), sino también para todos los problemas duales de programación lineal. Utilicemos ese teorema en el p. 3.2.3 para argumentar el método simplex de los problemas de programación lineal de forma canónica.

3.2.3. Demostración del método simplex.

Proposición 1. *El punto x admisible en el problema* (Pr_{can}) *es el punto extremo del conjunto de elementos admisibles, si y sólo si las columnas de la matriz A, correspondientes a las coordenadas positivas del vector x, son linealmente independientes.*

◁ Demostremos la necesidad y suficiencia partiendo de lo contrario. A) Sea x un punto extremo del conjunto C. Supongamos que son linealmente dependientes las columnas de la matriz A que corresponden a las coordenadas positivas del vector x. Estimemos, para mayor precisión, que esas columnas son $a^1, \ldots, a^k$. De la dependencia lineal se desprende que habrá un conjunto no nulo de números $\lambda_1, \ldots, \lambda_k$ para el cual $\sum_{i=1}^{k} \lambda_i a^i = 0$, es decir, $A\lambda = 0$ para el vector $\lambda = (\lambda_1, \ldots, \lambda_k, 0, \ldots, 0) \in \mathbf{R}^n$. Entonces el vector $x(\theta) = x + \theta\lambda$ será admisible para θ próximos a cero, lo que contradice el hecho de que x es el punto extremo del conjunto de elementos admisibles.

B) Supongamos que el punto admisible x y las columnas de la matriz A que corresponden a las coordenadas positivas del vector x son linealmente independientes. Para mayor precisión otra vez consideraremos que dichas columnas son $a^1, \ldots, a^k$ y que x no es el punto extremo del conjunto C. Entonces se hallarán ciertos puntos admisibles y y z, $y \neq z$, distintos de x, tales que, para cierto α, $0 < \alpha < 1$, $x = \alpha\, y + (1 - \alpha)\, z$. De esta última igualdad se deduce que no serán nulas las coordenadas de los puntos y y z que corresponden a las coordenadas de x, es decir, las primeras k coordenadas. Dado que $\sum_{i=1}^{k} y_i a^i = b$, $\sum_{i=1}^{k} z_i a^i = b$, entonces $\sum_{i=1}^{k} (y_i - z_i)\, a^i = 0$. De aquí se deduce que los vectores $a^1, \ldots, a^k$ son linealmente dependientes, lo que contradice la condición. Por consiguiente, nuestra suposición es incorrecta y el punto x es el punto extremo. ▷

Proposición 2. *Cualquier punto de un problema no degenerado posee no menos de m coordenadas positivas.*

◁ Demostrémoslo de lo contrario. Supongamos que existe un punto admisible con menos de m coordenadas positivas. Para mayor precisión considereramos que aquí se trata de las primeras k coordenadas ($k < m$). Analicemos el conjunto $\overline{C}$ de soluciones del sistema $\overline{A}y = b$, $y \geqslant 0$, donde $y = (y_1, \ldots, y_k) \in \mathbf{R}^k$, $\overline{A} = \{a^1, \ldots, a^k\}$ es una matriz del orden de $k \times n$. Sea $\hat{y}$ el punto extremo del conjunto $\overline{C}$ (se entiende bien que ese punto existe). Entonces el punto $\hat{x} = (\hat{y}, 0) \in \mathbf{R}^n$, siendo extremo en el conjunto C, tendrá menos de m coordenadas positivas, lo cual contradice el carácter no degenerado del problema. $\triangleright$

Proposición 3. *Un punto admisible en un problema no degenerado contiene exactamente m coordenadas positivas si y sólo si el mismo es el punto extremo de un conjunto de puntos admisibles.*

$\triangleleft$ Demostremos la necesidad partiendo de lo contrario. Supongamos que el punto admisible en un problema no degenerado contiene m coordenadas positivas (para mayor precisión supongamos que son las primeras). Además, supongamos que ese punto no es el punto extremo. Entonces, según la proposición 1, las columnas $a^1, \ldots, a^m$ de la matriz A serán linealmente dependientes. Eso significa que entre las soluciones del sistema $\overline{A}y = b$, $y \geqslant 0$, donde $\overline{A} = \{a^1, \ldots, a^m\}$, $y = (y_1, \ldots, y_m) \in \mathbf{R}^m$, habrá una solución y que poseerá menos de m coordenadas positivas. Por consiguiente, el vector $x = (y, 0) \in \mathbf{R}^n$ es el punto admisible (es decir, $x \in C$) que tiene menos de m coordenadas positivas, lo cual contradice la proposición 2.

La suficiencia se deduce de la definición del problema no degenerado. $\triangleright$

Teorema. *Sea* $\hat{x} = (x_1, \ldots, x_m, 0, \ldots, 0) \in \mathbf{R}^n$, $\hat{x}_i > 0$, $i = 1, \ldots, m$ *el punto extremo en el problema no degenerado* ($\mathrm{Pr_{can}}$) *de programación lineal en forma canónica. Entonces:*

a) *si* $\Delta \geqslant 0$, $\hat{x}$ *será la solución del problema;*

b) *si* $\Delta_j < 0$ *y* $x^j \leqslant 0$ *para cierto* j, *el valor del problema será* $S_{\text{máx}} = +\infty$;

c) *si no se cumplen las condiciones de los puntos* a) *y* b), x_{nuevo} *será el punto extremo del conjunto de elementos admisibles; en este caso el valor de la funcional aumentará en* $-\theta_0\Delta_{j_0}$ *y el desarrollo de los vectores* x_{nuevo}, $a^1, \ldots, a^n$ *se realizará por el método simplex.*

$\triangle$ a) Denotemos $y = A_{\text{bás}}^{*-1} c_{\text{bás}}$ (y recordemos que $\det A_{\text{bás}} \neq 0$). En vista de que $j = m+1, \ldots, n$,

$$\Delta_j = \langle c_{\text{bás}}, x^j \rangle - c_j = \langle A_{\text{bás}}^* y, x^j \rangle - c_j = \langle y, A_{\text{bás}} x^j \rangle - c_j = \\ = \langle y, a^j \rangle - c_j \geqslant 0$$

eso significa que $A_{\text{no bás}}^* y \geqslant c_{\text{no bás}}$ y, por lo tanto, $A^* y \geqslant c$, o sea, y es el elemento admisible en el problema (Pr) del p. 3.2.2, el cual es dual respecto al problema inicial ($\mathrm{Pr_{can}}$) de forma canónica (p. 3.1.1). Por otra parte, dado que las primeras m coordenadas del vector $A^* y - c$ son nulas, resulta que

$$\langle A^* y - c, \hat{x} \rangle = 0 \Leftrightarrow \langle c, \hat{x} \rangle = \langle b, y \rangle.$$

Así pues, según el criterio de solución, $\hat{x}$ es precisamente la solución del problema.

b) Supongamos que $x(t) = \hat{x} - tx^j + te_j$. Entonces $x(t) \geqslant 0$ y

$$Ax(t) = A\hat{x} - tA_{\text{bás}}x^j + tAe_j = A\hat{x} - ta^j + ta^j = b,$$

o sea, $x(t)$ es un elemento admisible $\forall t \geqslant 0$ y, por lo tanto,

$$\langle c, x(t)\rangle - \langle c, \hat{x}\rangle = -t\langle c, x^j\rangle + t\langle c, e_j\rangle = -t(z_j - c_j) = \\ = t\Delta_j \to +\infty)$$

cuando $t \to +\infty$.

c) Supongamos que no se cumplen las condiciones de los puntos a) y b) del teorema. Entonces, para cierto $m + 1 \leqslant j_0 \leqslant n$, $\Delta_{j_0} < 0$ y $\theta_0 = \min\{x_i/x_{ij_0} \mid x_{ij_0} > 0\} = x_{i_0}/x_{i_0j_0} > 0$. Tomamos $l = (l_{\text{bás}}, l_{\text{no bás}})$, $l_{\text{no bás}} = (0, \ldots, 1, \ldots, 0)$, $l_{\text{bás}} = -A^{-1}_{\text{bás}}A_{\text{no bás}} \times l_{\text{no bás}}$, $x(\theta_0) = \hat{x} - \theta_0 l$. Entonces $As(\theta_0) = A\hat{x} + \theta_0 Al = A\hat{x} = b$, $x_{\text{no bás}}(\theta_0) = \theta_0 l_{\text{no bás}} \geqslant 0$, $x_{\text{bás}}(\theta_0) = x_{\text{bás}} + \theta_0 l_{\text{bás}} = x_{\text{bás}} - \theta_0 A^{-1}_{\text{bás}}A_{\text{no bás}}l_{\text{no bás}} = x_{\text{bás}} - \theta_0 A^{-1}_{\text{bás}}a^{j_0} = x_{\text{bás}} - \theta_0 x^{j_0} \geqslant 0$ en virtud de la elección de θ_0, es decir, $x(\theta_0)$ es el punto admisible. En vista de que, según la proposición 2, el punto admisible en un problema no degenerado contiene no menos de m coordenadas positivas, tan sólo una coordenada x_{i_0} del vector $x_{\text{bás}} - \theta_0 x^{j_0}$ se convertirá en cero. Consiguientemente, el punto admisible $x_{\text{bás}}(\theta_0)$ tiene exactamente m coordenadas positivas en los lugares $1, \ldots, i_0 - 1, i_0 + 1, \ldots, m, j_0$ y, por lo tanto, según la proposición 3, el punto $x_{\text{nuevo}} = x(\theta_0)$ será el punto extremo. Además, el valor de la funcional aumentará en $-\theta_0\Delta_{j_0}$ (véase la demostración en el p. b).

Las fórmulas $(x_{\text{nuevo}})_i = x_i - \theta_0 x_{ij_0} = x_i - x_{ij_0}x_{i_0}/x_{i_0j_0}$, $i = 1, \ldots, m$, $(x_{\text{nuevo}})_{j_0} = \theta_0 = x_{i_0}/x_{i_0j_0}$ significan que la columna x_{nueva} en la nueva tabla simplex ha de escribirse según el referido método de confección de dicha tabla. Mostremos que las demás columnas $(x_{\text{nueva}})^j$ también se construyen con arreglo a ese mismo método. Para este fin calculemos las coordenadas $(x_{ij})_{\text{nueva}}$ de desarrollo de los vectores a^j, $j = 1, \ldots, n$ según la base $a^1, \ldots, a^{i_0-1}, a^j, a^{i_0+1}, \ldots, a^m$.

En la base anterior

$$a^j = \sum_{i=1}^{m} a^i x_{ij} = \sum_{\substack{i=1 \\ i \neq i_0}}^{m} a^i x_{ij} + a^{i_0}x_{i_0j},$$

$$j = 1, \ldots, n,$$

de aquí, para $j = j_0$ obtenemos

$$a^{j_0} = \sum_{\substack{i=1 \\ i \neq i_0}}^{m} a^i x_{ij_0} + a^{i_0}x_{i_0j_0}.$$

Como $x_{i_0j_0} > 0$, resulta que

$$a^{i_0} = -\sum_{\substack{i=1\\ i\neq i_0}}^{m} a^i \frac{x_{ij_0}}{x_{i_0j_0}} + \frac{a^{j_0}}{x_{i_0j_0}},$$

entonces

$$a^j = \sum_{\substack{i=1\\ i\neq i_0}}^{m} \left(x_{ij} - \frac{x_{ij_0}x_{i_0j}}{x_{i_0j_0}}\right) a^i + \frac{x_{i_0j}}{x_{i_0j_0}} a^{j_0}, \quad j=1, \ldots, n.$$

Por consiguiente,

$$(x_{\text{nuevo}})_{i_0j} = \frac{x_{i_0j}}{x_{i_0j_0}}, \quad j=1, \ldots, n \Rightarrow (x_{\text{nuevo}})_{i_0j_0} = 1,$$

$$(x_{\text{nuevo}})_{ij} = x_{ij} - x_{ij_0}x_{i_0j}/x_{i_0j_0},$$

$$i \neq i_0, \ i=1, \ldots, m, \ j=1, \ldots, n \Rightarrow (x_{ij_0})_{\text{nuevo}} = 0, \ i \neq i_0. \ \triangleright$$

§ 4. Cálculo clásico de variaciones

En este párrafo se deducen las condiciones necesarias de primer orden para algunas clases de problemas que tradicionalmente se examinan en el cálculo clásico de variaciones (c.c.v.): problemas de Bolz, problemas elementales de c.c.v., problemas isométricos, etc. Al deducir las condiciones necesarias de extremo, hemos tratado de utilizar tan sólo los medios elementales del análisis clásico, por eso los materiales de este párrafo pueden ser estudiados independientemente del párrafo anterior.

4.1. Problema de Bolz

4.1.1. Planteamiento del problema. Llámase *problema de Bolz* el siguiente problema extremal de funciones continuamente diferenciables $C^1([t_0, t_1])$ no acotadas en el espacio:

$$\mathscr{B}(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt + l(x(t_0), y(t_1)) \to \text{extr.} \qquad \text{(Pr)}$$

Aquí $L = L(t, x, \dot{x})$ es una función de tres variables; y $l = l(x_0, x_1)$, una función de dos variables. El segmento $[t_0, x_1]$ se considera fijo y finito, $-\infty < t_0 < t_1 < +\infty$. El problema de Bolz es un problema elemental del cálculo clásico de variaciones.

La función L se llama integrante; la función l, terminente; y la funcional $\mathscr{B}$, funcional de Bolz.

Digamos que la función $\hat{x}(\cdot) \in C^1([t_0, t_1])$ proporciona al problema (Pr) un mínimo (máximo) local (débil) o, que es lo mismo, a la funcional $\mathscr{B}$ en el espacio $C^1([t_0, t_1])$, y, escribimos $\hat{x}(\cdot) \in \in \text{loc mín Pr}$ (loc máx Pr) si se halla $\delta > 0$ tal que con cualquier función $x(\cdot) \in C^1([t_0, t_1])$ para la cual $\| x(\cdot) - \hat{x}(\cdot) \|_1 < \delta \times \times (\Leftrightarrow \| x(\cdot) - \hat{x}(\cdot) \|_{C([t_0, t_1])} < \delta, \ \| \dot{x}(\cdot) - \dot{\hat{x}}(\cdot) \|_{C([t_0, t_1])} < \delta$ se cumple la desigualdad

$$\mathscr{B}(x(\cdot)) \geqslant \mathscr{B}(\hat{x}(\cdot)) \ (\mathscr{B}(x(\cdot)) \leqslant \mathscr{B}(\hat{x}(\cdot))).$$

A la par con el extremo débil, en el cálculo clásico de variaciones tradicionalmente también se estudia el extremo fuerte. En este caso se amplía algo la clase de funciones en las que se examina la funcional $\mathscr{B}$. El extremo en el problema se busca entre las funciones $x(\cdot)$ pertenecientes a $KC^1([t_0, t_1])$, es decir, entre las funciones continuamente diferenciables a trozos. En el p. 9.2.1 regresaremos al concepto de extremo fuerte.

Al resolver problemas del cálculo de variaciones utilizaremos de aquí en adelante, los términos de «extremo absoluto» o «extremo global». A estos términos se les puede dar un sentido corriente, según el cual la función hallada tiene valor extremal de la funcional dentro de todas las funciones admisibles (en nuestro caso las funciones admisibles pertenecen a C^1 o a KC^1). No obstante, por regla general, las funciones que proporcionan el extremo absoluto a C^1 o a KC^1, en realidad proporcionan ese extremo a una clase más amplia de funciones, mejor dicho, a todas las funciones absolutamente continuas en las que ha sido definida la funcional.

4.1.2. Condiciones necesarias de extremo.

Teorema. *Supongamos que la función $\hat{x}(\cdot) \in C^1([t_0, t_1])$ proporciona un extremo local débil al problema de Bolz* (Pr). *Se estima que el integrante L es continuo junto con sus derivadas parciales en x y $\dot{x}$ en cierto entorno del conjunto $\{(t, \hat{x}(t), \dot{\hat{x}}(t)) \mid t \in [t_0, t_1]\}$, mientras que el terminante l es continuamente diferenciable en el entorno del punto $(\hat{x}(t_0), \hat{x}(t_1))$.*

Entonces $\hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$ $(\hat{L}_{\dot{x}(t)} = L_{\dot{x}}(t, \hat{x}(t), \dot{\hat{x}}(t)))$ y son cumplidas

a) *la ecuación de Euler*

$$-\frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0;$$

b) *las condiciones de transversalidad*

$$\hat{l}_{x_k} = l_{x_k}(\hat{x}(t_0), \hat{x}(t_1))$$

$$\hat{L}_{\dot{x}}(t_k) = (-1)^k \hat{l}_{x_k}, \ k = 0, 1.$$

◁ Demostremos el teorema en varias etapas.

A) **Definición de la variación de Lagrange.** Tomemos una función $x(\cdot) \in C^1([t_0, t_1])$ aleatoria pero fija. Dado que $\hat{x}(\cdot) \in \operatorname{loc\,extr}$ Pr, la función de una variable

$$\varphi(\lambda) := \mathscr{B}(\hat{x}(\cdot) + \lambda x(\cdot)) = \int_{t_1}^{t_0} L(t, \hat{x}(t) + \lambda x(t), \dot{\hat{x}}(t) + \lambda \dot{x}(t)\, dt +$$

$$+ l(\hat{x}(t_0) + \lambda x(t_0), \hat{x}(t_1) - \lambda x(t_1))$$

tendrá extremo cuando $\lambda = 0$. Supongamos que $F(t, \lambda) - L(t, \hat{x}(t) + \lambda x(t), \dot{\hat{x}}(t) + \lambda \dot{x}(t))$. De las condiciones de suavidad aplicadas a L, $\hat{x}(\cdot)$, $x(\cdot)$ se deduce que la función $\varphi(\lambda)$ es diferenciable en cero. Efectivamente, las funciones F y F_λ son continuas en cierto rectángulo $[t_0 t_1] [-\lambda_0, \lambda_0]$ y, por lo tanto, valiéndonos del conocido teorema del análisis, podemos efectuar la diferenciación bajo el signo de integral (15, t. 2, pág. 107). Pero entonces, según el teorema de Fermat, $\varphi'(0) = 0$. Diferenciando la función φ y suponiendo que $\lambda = 0$, obtenemos

$$\varphi'(0) = \lim_{\lambda \to 0} \frac{\mathscr{B}(\hat{x}(\cdot) + \lambda x(\cdot)) - \mathscr{B}(\hat{x}(\cdot))}{\lambda} \overset{\text{def}}{=} \delta\mathscr{B}(\hat{x}(\cdot), x(\cdot)) =$$

$$= \int_{t_0}^{t_1} (\hat{L}_x(t)\, x(t) + \hat{L}_{\dot{x}}(t)\, \dot{x}(t))\, dt + \hat{l}_{x_0} x(t_0) + \hat{l}_{x_1}(t_1) = 0 \qquad (1)$$

$$\forall x(\cdot) \to C^1([t_0, t_1]).$$

Así pues, hemos calculado la variación de Lagrange según la funcional de Bolz $\mathscr{B}$ y hemos aclarado que la condición necesaria para un extremo local débil de dicha funcional en $\hat{x}(\cdot)$ consiste en la igualdad a cero de su variación de Lagrange.

B) **Lema de Dubois — Reimond.** *Sean continuas las funciones $a_0(\cdot)$ y $a_1(\cdot)$ en el segmento $[t_0, t_1]$ y sea cumplida, para cualquier función continuamente diferenciable $x(\cdot)$, $x(t_0) = x(t_1) = 0$, la igualdad*

$$\int_{t_0}^{t_1} (a_1(t)\, \dot{x}(t) + a_0(t)\, x(t))\, dt = 0.$$

Entonces la función $a_1(\cdot)$ será continuamente diferenciable y

$$-\frac{da_1(t)}{dt} + a_0(t) = 0.$$

$\lhd\lhd$ Tomamos una función $p(\cdot) \in C^1([t_0, t_1])$ tal que $\dot{p}(t) = a_0(t)$ e $\int_{t_0}^{t_1} p(t)\, dt = \int_{t_0}^{t_1} a_1(t)\, dt$. Entonces, para cualquier función $x(\cdot) \in C^1([t_0, t_1)]$, $x(t_0) = x(t_1) = 0$ deberá cumplirse, según la condición del lema, la igualdad

$$0 = \int_{t_0}^{t_1} (a_1(t)\dot{x}(t) + a_0(t)x(t))\,dt = \tag{2}$$

$$= \int_{t_0}^{t_1} a_1(t)\dot{x}(t)\,dt + \int_{t_0}^{t_1} x(t)\,dp(t) = \int_{t_0}^{t_1} (a_1(t) - p(t)\dot{x}(t)\,dt.$$

Elegimos una función $\tilde{x}(\cdot) \in C^1([t_0, t_1)$ tal que $\dot{\tilde{x}}(t) = a_1(t) - p(t)$, $\tilde{x}(t_0) = 0$, Entonces, en virtud de la elección de la función $p(\cdot)$,

$$\tilde{x}(t_1) = \int_{t_0}^{t_1} \dot{\tilde{x}}(t)\,dt = \int_{t_0}^{t_1} (a_1(t) - p(t))\,dt = 0.$$

Por consiguiente, para la función $x(\cdot) = \tilde{x}(\cdot)$ ha de cumplirse la igualdad (2), o sea, $\int_{t_0}^{t_1} (a_1(t) - p(t))^2\,dt = 0$. De esta última relación se deduce que $a_1(t) \equiv p(t)$, es decir, $a_1(\cdot) \in C^1([t_0, t_1])$, $-\frac{da_1(t)}{dt} + a_0(t) = 0$. $\triangleright\triangleright$

C) **Terminación de la demostración.** La igualdad (1) se cumple para cualquier función $x(\cdot) \in C^1([t_0, t_1])$ y, por lo tanto, para todas las funciones $x(\cdot) \in C_0^1(t_0, t_1) = \{x(\cdot) \in C^1([t_0, t_1]) \mid x(t_0) = x(t_1) = 0\}$. Por consiguiente, de (1) se deduce que

$$\int_{t_0}^{t_1} (\hat{L}_{\dot{x}}(t)\dot{x}(t) + \hat{L}_x(t)x(t))\,dt = 0$$

$$\forall x(\cdot)\, C_0^1([t_0, t_1]).$$

Según el lema de Dubois — Reimond, (3)

$$\hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1]) \text{ y}$$

$$\frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0.$$

Integrando por partes en la igualdad 1 (lo cual se ha hecho posible en virtud de la inclusión demostrada $\hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$) y teniendo en cuenta (3), obtenemos

$$\int_{t_0}^{t_1} (\hat{L}_x(t)\,x(t) + \hat{L}_{\dot{x}}(t)\,\dot{x}(t))\,dt + \hat{l}_{x_0}x(t_0) + \hat{l}_{x_1}x(t_1) =$$

$$= \int_{t_0}^{t_1} \left(\hat{L}_x(t) - \frac{d}{dt}\hat{L}_{\dot{x}}(t)\right) x(t)\,dt + x(t)\,\hat{L}_{\dot{x}}(t)\Big|_{t_0}^{t_1} + \hat{l}_{x_0}x(t_0) + \hat{l}_{x_1}x(t_1) =$$

$$= (\hat{L}_{\dot{x}}(t_1) + \hat{l}_{x_1})\,x(t_1) + (-\hat{L}_{\dot{x}}(t_0) + \hat{l}_{x_0})\,x(t_0) = 0$$

$$\forall x(\cdot) \in C^1([t_0, t_1]). \tag{4}$$

Introduciendo sucesivamente $x(t) = t - t_0$ y $x(t) = t - t_1$ en la expresión (4), llegaremos a las condiciones de transversalidad $\hat{L}_{\dot{x}}(t_1) = -\hat{l}_{x_1}$ y $\hat{L}_{\dot{x}}(t_0) = \hat{l}_{x_0}$. ▷

Nota. Debemos señalar que esas tres etapas de demostración del teorema aparecerán de una u otra forma al demostrar otros teoremas del cálculo clásico de variaciones y de control óptimo.

El conjunto de condiciones para hallar el extremo local débil, es un conjunto completo. Efectivamente, la ecuación de Euler es una ecuación diferencial de segundo orden. Su solución general comprende dos constantes incógnitas. Para determinarlas existen dos ecuaciones: las condiciones de transversalidad.

Hemos formulado el teorema para el problema unidimensional de Bolz. Señalemos ciertas modificaciones necesarias que deben introducirse para el caso vectorial.

Supongamos que en el problema (Pr), $x(\cdot) = (x_1(\cdot), \ldots, x_n(\cdot))$, $L = L(t, x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)$ es una función de $2n + 1$ variables; y $l = l(x_{01}, \ldots, x_{0n}, x_{11}, \ldots, x_{1n})$, una función de $2n$ variables. Las condiciones necesarias en el problema vectorial de Boltz constan del sistema de ecuaciones de Euler

$$-\frac{d}{dt}\hat{L}_{\dot{x}_i}(t) + \hat{L}_{x_i}(t) = 0, \quad i = 1, \ldots, n,$$

y de las condiciones de transversalidad establecidas por el sistema de ecuaciones

$$\hat{L}_{\dot{x}_i}(t_k) = (-1)^k\,\hat{l}_{x_{ki}}(t_k), \; i = 1, \ldots, n, \; k = 0, 1.$$

La demostración del teorema en el caso vectorial se reduce trivialmente a una demostración unidimensional.

4.1.3. Ejemplo.

$$\mathscr{B}(x(\cdot)) = \int_0^1 (\dot{x}^2 - x)\,dt + x^2(1) \to \text{extr}.$$

Condiciones necesarias:

a) ecuación de Euler $-\frac{d}{dt} L_{\dot{x}} - L_x = 0 \Leftrightarrow 2\ddot{x} + 1 = 0$;

b) condiciones de transversalidad

$$L_{\dot{x}}(0) = l_{x(0)},\ L_{\dot{x}}(1) = -l_{x(1)} \Leftrightarrow \dot{x}(0)$$

$$\dot{x}(1) = -x(1).$$

Solución general de la ecuación euleriana: $x(t) = -\frac{t^2}{4} + C_1 t + C_2$. Valiéndonos de las condiciones de transversalidad hallamos que $C_1 = 0$, $C_2 = 3/4$. En resumen, existe la única extremal admisible $x(t) = (3 - t^2)/4$.

Mostremos que esa extremal proporciona un mínimo absoluto al problema. Efectivamente, si $h(\cdot) \in C^1([t_0 t_1])$, entonces

$$\mathscr{B}(\hat{x}(\cdot) + h(\cdot)) - \mathscr{B}(\hat{x}(\cdot)) = \int_0^1 2\dot{\hat{x}}\dot{h}\,dt + \int_0^1 \dot{h}^2\,dt -$$

$$-\int_0^1 h\,dt + 2\hat{x}(1)h(1) + h^2(1).$$

Integrando por partes y teniendo en cuenta que $\hat{x}(t) = (3 - t^2)/4$, obtenemos

$$\mathscr{B}(\hat{x}(\cdot) + h(\cdot)) - \mathscr{B}(\hat{x}(\cdot)) = 2\dot{\hat{x}}h\Big|_0^1 - \int_0^1 (2\ddot{\hat{x}} + 1)h\,dt +$$

$$+\int_0^1 \dot{h}^2\,dt + 2\hat{x}(1)h(1) + h^2(1) = \int_0^1 \dot{h}^2\,dt + h^2(1) \geqslant 0.$$

Resultado. $\hat{x}(t) - (3 - t^2)/4 \in \text{absmín}$, $S_{\text{máx}} = +\infty$.

4.2. Problema elemental del cálculo clásico de variaciones

4.2.1. Planteamiento del problema. Llámase problema elemental *del cálculo clásico de variaciones*, el siguiente problema extremal en $C^1([t_0, t_1])$:

$$\mathcal{J}(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\,dt \to \text{extr};$$

$$x(t_0) = x_0,\ x(t_1) = x_1. \qquad \text{(Pr)}$$

Aquí $L = L(t, x, \dot{x})$ es una función de tres variables, denominada *integrante*. El extremo en el problema se halla entre las funciones

$x(\cdot) \in C^1([t_0, t_1])$ que satisfacen *las condiciones en los extremos o en las condiciones de contorno* $x(t_0) = x_0$, $x(t_1) = x_1$. Tales funciones se llaman *admisibles*.

Decimos que la función admisble $\hat{x}(\cdot)$ proporciona un *mínimo (máximo) local débil* al problema (Pr), y escribimos $\hat{x}(\cdot) \in$ locmín Pr (locmáx Pr) si existe $\delta > 0$ tal que, con cualquier función admisible $x(\cdot)$ para la cual $\| x(\cdot) - \hat{x}(\cdot) \|_1 < \delta$, se cumpla la desigualdad

$$\mathcal{J}(x(\cdot)) \geqslant \mathcal{J}(\hat{x}(\cdot)) \; (\mathcal{J}(x(\cdot)) \leqslant \mathcal{J}(\hat{x}(\cdot))).$$

4.2.2. Condición necesaria de extremo.

Teorema. *Supongamos* que $\hat{x}(\cdot) \in C^1([t_0, t_1])$ *proporciona un extremo local débil a un problema elemental del cálculo clásico de variaciones, y el integrante L es continuo juntamente con sus derivadas parciales en x y $\dot{x}$ en cierto entorno del conjunto* $\{(t, \hat{x}(t), \hat{\dot{x}}(t)) \mid t \in [t_0, t_1]\}$. *Entonces $\hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$ y se cumple la ecuación de Euler*

$$-\frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0.$$

◁ Razonamos igualmente que hemos razonado al deducir las condiciones necesarias en el problema de Bolz. Tomamos la función aleatoria, pero fija, $x(\cdot) \in C_0^1([t_0, t_1])$. En este caso $\hat{x}(\cdot) + \lambda x(\cdot)$ será la función admisible $\forall \lambda \in \mathbf{R}$. Supongamos que $\varphi(\lambda) = \mathcal{J}(\hat{x}(\cdot) + \lambda x(\cdot))$. De la condición $\hat{x}(\cdot) \in$ locextr Pr se deduce que $0 \in$ locextr φ. Utilizando la diferenciabilidad de la función φ en cero y la expresión para la variación de la funcional (p. 4.1.2), obtenemos

$$\varphi'(0) = \delta\mathcal{J}(\hat{x}(\cdot), x(\cdot)) = \int_{t_0}^{t_1} (\hat{L}_{\dot{x}}(t)\dot{x}(t) + \hat{L}_x(t)x(t))\,dt = 0$$

$$\forall x(\cdot) \in C_0^1([t_0, t_1]).$$

Del lema Dubois — Reimond se deduce que $\hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$ y que, por lo tanto, se cumple la ecuación de Euler. ▷

El conjunto de condiciones para hallar la extremal admisible es un conjunto completo. Además, la ecuación de Euler es una ecuación diferencial de segundo orden. Su solución general contiene dos constantes incógnitas. Para determinarlas existen dos ecuaciones: las llamadas condiciones en los extremos. Por consiguiente, la extremal más a menudo admitida es una extremal única.

Hemos formulado el teorema para un problema elemental unidimensional del cálculo clásico de variaciones. Señalemos los cambios necesarios que requiere el caso vectorial.

Sea $x(\cdot) = (x_1(\cdot), \ldots, x_n(\cdot))$, $L = L(t, x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)$ una función de $2n + 1$ variables en el problema (Pr) del

p. 4.2.1. Las condiciones necesarias en un problema vectorial elemental constan del sistema de ecuaciones de Euler

$$-\frac{d}{dt}\hat{L}_{\dot{x}_i}(t)+\hat{L}_{x_i}(t)=0, \quad i=1, \ldots, n.$$

La demostración del teorema en el caso vectorial trivialmente se reduce a una demostración unidimensional.

4.2.3. Integrales de la ecuación de Euler. Si el integrante $L = L(t, x, \dot{x})$ no depende explícitamente de una de las variables, la ecuación euleriana se reducirá a una ecuación más simple.

1. Si el integrante $L = L(t, x)$ no depende explícitamente de $\dot{x}$, la ecuación euleriana se reducirá a la ecuación

$$\hat{L}_x(t) = 0.$$

2. Si el integrante $L = L(t, \dot{x})$ no depende explícitamente de x, tendrá lugar la integral de impulso

$$\hat{L}_{\dot{x}}(t) = \text{const}.$$

3. Si el integrante $L = L(x, \dot{x})$ no depende explícitamente de t, tendrá lugar la integral de energía (ambas denominaciones de dichos términos se han tomado de la mecánica clásica)

$$\hat{\dot{x}}\hat{L}_{\dot{x}}(t) - \hat{L}(t) = \text{const}$$

(para la demostración es suficiente diferenciar la relación con arreglo a t y utilizar la ecuación de Euler).

4.2.4. Ejemplos. En este punto, mediante ejemplos examinaremos las relaciones entre las soluciones de un problema elemental del cálculo clásico de variaciones y las extremales.

Ejemplo 1 (la extremal admisible existe, es única y proporciona un extremo global).

$$\mathcal{J}(x(\cdot))=\int_0^1 \dot{x}^2\,dt \to \inf; \quad x(0)=0, \quad x(1)=1.$$

Ecuación de Euler: $\ddot{x} = 0$.

Solución general: $x = C_1t + C_2$. La única extremal admisible es $\hat{x} = t$.

La extremal proporciona un mínimo global al problema. Efectivamente, sea $h(\cdot) \in C_0^1([0, 1])$. Entonces $\hat{x}(\cdot) + h(\cdot)$ será una función admisible en el problema y

$$\mathcal{J}(\hat{x}(\cdot)+h(\cdot))-\mathcal{J}(\hat{x}(\cdot))=\int_0^1 (\dot{\hat{x}}+\dot{h})^2\,dt-\int_0^1 \dot{\hat{x}}\,dt=$$

$$=2\int_0^1 \dot{h}\,dt+\int_0^1 \dot{h}^2\,dt=\int_0^1 \dot{h}^2\,dt \geqslant 0.$$

En ese ejemplo todo está bien, pero en los que siguen surgen complicaciones diferentes.

Ejemplo 2 (la extremal existe, es única, proporciona un extremo global, pero no es una función continuamente diferenciable).

$$\mathcal{J}(x(\cdot))=\int_0^1 t^{2/3}\dot{x}^2\,dt \to \inf;\quad x(0)=0,\quad x(1)=1$$

(ejemplo de Hilbert).

Ecuación de Euler: $\frac{d}{dt}(2t^{2/3}\dot{x})=0 \Leftrightarrow t^{2/3}\dot{x}=C \Leftrightarrow \dot{x}=Ct^{-2/3}$,

Solución general: $x=C_1t^{1/3}+C_2$. La única extremal que satisface las condiciones en los extremos es $\hat{x}=x^{1/3}$.

Está claro que la extremal no es una función de clase $C^1([0,1])$, ya que $\hat{x}(\cdot)\in C([0,1])$. Mostremos, no obstante, que ella proporciona un mínimo global al problema entre todas las funciones absolutamente continuas $x(\cdot)$ que satisfacen las condiciones de contorno y para las cuales la integral $\mathcal{J}$ es finita. Efectivamente, para cualquier $h(\cdot)$ de este tipo,

$$\mathcal{J}(\hat{x}(\cdot)+h(\cdot))-\mathcal{J}(\hat{x}))=\int_0^1 t^{2/3}\left(\frac{t^{-2/3}}{3}+\dot{h}\right)^2\times$$

$$\times\,dt-\int_0^1 t^{2/3}\left(\frac{t^{-2/3}}{3}\right)^2 dt=\int_0^1 \frac{2}{3}\dot{h}\,dt+\int_0^1 t^{2/3}\dot{h}^2\,dt\geqslant 0.$$

Ejemplo 3 (no existe solución del problema ni extremal admisible incluso entre las funciones absolutamente continuas).

$$\mathcal{J}(x(\cdot))=\int_0^1 t^2\dot{x}^2\,dt\to\inf;\quad x(0)=0,\quad x(1)=1$$

(ejemplo de Weierstrass).

Ecuación de Euler: $\frac{d}{dt}(2t^2\dot{x})=0\Leftrightarrow t^2\dot{x}=C\Leftrightarrow \dot{x}=C/t^2$.

Solución general: $x=(C_1/t)+C_2$.

No hay ninguna extremal que satisfaga la condición de contorno $x(0)=0$

Es obvio que $\mathcal{J}(x(\cdot))\geqslant 0$, y para cualquier función absolutamente continua $(x(\cdot)\not\equiv 0$ $\mathcal{J}(x(\cdot))>0$. Mostremos que la cota inferior en el problema es igual a cero. Analicemos la sucesión de las funciones admisibles $x_n(t)=\operatorname{arctg} nt/\operatorname{arctg} n$. Tenemos

$$\mathcal{J}(x_n(\cdot))=\int_0^1 t^2\frac{n^2}{(1+n^2t^2)\operatorname{arctg}^2 n}\,dt\leqslant$$

$$\leqslant\int_0^{1/n}\frac{dt}{\operatorname{arctg}^2 n}+\int_{1/n}^1\frac{dt}{n^2t^2\operatorname{arctg}^2 n}\to 0.$$

Ejemplo 4 (la extremal admisible existe, es única, pero no proporciona extremo).

$$\mathcal{J}(x(\cdot))=\int_0^{3\pi/2}(\dot{x}^2-x^2)\,dt\to\inf;\quad x(0)=x\left(\frac{3\pi}{2}\right)=0.$$

Ecuación de Euler: $\ddot{x} + x = 0$.

Solución general: $x = C_1 \operatorname{sen} t + C_2 \cos t$. La única extremal admisible es $\hat{x} \equiv 0$.

Examinemos la sucesión de funciones $x_n(t) = (1/n)\cdot \operatorname{sen}(2t/3)$. Es obvio que $x_n(\cdot)$ son funciones admisibles y $x_n(\cdot) \to x(\cdot) \equiv 0$ en $C^1([0, 3\pi/2])$, pero en este caso

$$\mathcal{J}(x_n(\cdot)) = \frac{1}{n^2}\,\frac{3\pi}{4}\left(\frac{4}{9} - 1\right) = -\frac{5\pi}{12n^2} < 0 = \mathcal{J}(\hat{x}(\cdot)).$$

Del ejemplo 4 se deduce, en particular, que la ecuación de Euler es una condición de extremo necesaria, pero no suficiente.

Ejemplo 5.

$$\mathcal{J}(x(\cdot)) = \int_0^1 ((1 - \dot{x}^2)^2 + x^2)\, dt \to \inf;$$

$$x(0) = 0, \quad x(1) = 0.$$

La cota inferior de la funcional es igual a cero. Para cerciorarse de ello es suficiente analizar la sucesión minimizadora de las funciones de $KC^1([0, 1])$:

$$x_n(t) = \int_0^t \operatorname{sign} \operatorname{sen} 2\pi n\tau\, d\tau, \quad n = 1, 2, \ldots$$

Las funciones $x_n(\cdot)$ tienden a cero con uniformidad, mientras que $|\dot{x}_n(t)| = 1$, a excepción de un número finito de puntos, es decir, $\mathcal{J}(x_n(\cdot)) \to 0$. Por otra parte, si $x_0(t) = 0$, tendremos $\mathcal{J}(x_0(\cdot)) = 1$, pero si $x(t) \not\equiv 0$, entonces $\mathcal{J}(x(\cdot)) \geqslant \int_0^1 x^2\, dt > 0$. Por lo tanto, la cota inferior de la funcional no es alcanzada en ninguna función admisible.

Las sucesiones de funciones semejantes a $x_n(\cdot)$ se llaman «regímenes deslizantes», ya que el control en ellas (en este ejemplo se trata de una derivada) parece como si se deslizara entre ciertos valores (en este caso entre $+1$ y -1).

4.3. Problemas isoperimétricos

4.3.1. Planteamiento del problema. Llámase *problema isoperimétrico* (con extremos fijos) en el cálculo clásico de variaciones, el siguiente problema en el espacio $C^1([t_0, t_1])$:

$$\mathcal{J}_0(x(\cdot)) = \int_{t_0}^{t_1} f_0(t, x(t), \dot{x}(t))\, dt \to \text{extr}; \qquad \text{(Pr)}$$

$$\mathcal{J}_i(x(\cdot)) = \int_{t_0}^{t_1} f_i(t, x(t), \dot{x}(t))\, dt = \alpha_i, \quad i = 1, \ldots, m \qquad (1)$$

$$x(t_0) = x_0, \quad x(t_1) = x_1, \qquad (2)$$

donde $\alpha_1, \ldots, \alpha_m$ son números fijos dados.

Las acotaciones de tipo (1) se llaman *isoperimétricas*. Las funciones f_i, $i = 0, 1, \ldots, m$ se llaman *integrantes*. Las funciones $x(\cdot) \in C^1([t_0, t_1])$ que satisfacen las condiciones isoperimétricas (1) y las condiciones en los extremos (2) se denominan *admisibles*.

Diremos que la función admisible $\hat{x}(\cdot)$ proporciona al problema (Pr) *un mínimo local débil* y escribimos $\hat{x}(\cdot) \in$ locmín Pr (locmáx Pr) si existe $\delta > 0$ tal que para cualquier función admisible $x(\cdot)$, a la cual le corresponde $\| x(\cdot) - \hat{x}(\cdot) \|_{C^1([t_0, t_1])} < \delta$, se cumple la desigualdad

$$\mathcal{J}_0(x(\cdot)) \geqslant \mathcal{J}_0(\hat{x}(\cdot)) \; (\mathcal{J}_0 x(\cdot)) \leqslant \mathcal{J}_0(\hat{x}(\cdot))).$$

4.3.2. Condición de extremo necesaria.

Teorema. *Sea que la función $\hat{x}(\cdot)$ proporciona un extremo local débil al problema isoperimétrico* (Pr). *Además, las funciones f_i, $i = 0, 1, \ldots, m$ y sus derivadas parciales en x y $\dot{x}$ son continuas en cierto entorno del conjunto $\{(t, \hat{x}(t), \dot{\hat{x}}(t)) \mid t \in [t_0, t_1]\}$ (condición de suavidad).*

Entonces se hallarán los factores de Lagrange $\lambda_0, \lambda_1, \ldots, \lambda_m$ que no todos son iguales a cero y que para la lagrangiana

$$L = \sum_{i=0}^{m} \lambda_i f_i(t, x, \dot{x}) \quad \hat{L}_{\dot{x}}(\cdot) \in C^1([t_0, t_1])$$

se cumple la ecuación de Euler

$$-\frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0.$$

A) **Variaciones de Lagrange de las funciones $\mathcal{J}_i$.** Escribamos las variaciones de Lagrange de las funcionales $\mathcal{J}_i$ (véase el p. 4.1.3):

$$\delta\mathcal{J}_i(\hat{x}(\cdot), x(\cdot)) = \int_{t_0}^{t_1} (\hat{f}_{i\dot{x}}(t)\dot{x}(t) + \hat{f}_{ix}(t)\, dt,$$

$$i = 0, 1, \ldots, m.$$

$$x(\cdot) \in C_0^1([t_0, t_1]).$$

B) **Construcción de la aplicación de dimensión finita y formación de los casos degenerado y no degenerado.** Examinemos la siguiente aplicación lineal del espacio $C_0^1([t_0, t_1]) = \{x(\cdot) \in C^1[(t_0, t_1) \times \times \mid x(t_0) = x(t_1) = 0\}$ en el espacio $\mathbf{R}^{m+1}$:

$$Ax(\cdot) = (\delta\mathcal{J}_0(\hat{x}(\cdot), x(\cdot)), \quad \delta\mathcal{J}_1(\hat{x}(\cdot), x(\cdot)), \ldots$$
$$\ldots, \delta\mathcal{J}_m(\hat{x}(\cdot)\, x(\cdot))).$$

Son posibles dos casos:

a) A es la aplicación a una parte de $\mathbf{R}^{m+1}$ (caso degenerado);

b) A es la aplicación a todos los espacios $\mathbf{R}^{m+1}$, es decir, $\operatorname{Im} A = \mathbf{R}^{m+1}$ (caso no degenerado).

C) Demostración del teorema en el caso degenerado. La imagen de un espacio lineal en una aplicación lineal es, según se sabe, un subconjunto. Por lo tanto, en el caso degenerado Im A es un subconjunto propio en $\mathbf{R}^{m+1}$. Pero entonces se hallarán los números $\lambda_0, \lambda_1, \ldots, \lambda_m$ que no todos son iguales a cero y con los que

$$\sum_{i=0}^{m} \lambda_i z_i = 0 \quad \forall z = (z_0, z_1, \ldots, z_m) \in \operatorname{Im} A.$$

Si recordamos la definición del operador A y la expresión para $\delta\mathcal{J}_i(\hat{x}(\cdot), x(\cdot))$, obtenemos

$$\int_{t_0}^{t_1} \left(\sum_{i=0}^{m} \lambda_i (\hat{f}_{i\dot{x}}(t)\dot{x}(t) + \hat{f}_{ix}(t)x(t)) \right) dt = 0$$

$$\forall x(\cdot) \in C_0^1([t_0, t_1]).$$

Pero entonces, del lema Dubois—Reimond (p. 4.1.3) se deduce que

$$\sum_{i=0}^{m} \lambda_i \hat{f}_{i\dot{x}}(\cdot) \in C^1([t_0, t_1]) \text{ y } -\frac{d}{dt}\left(\sum_{i=0}^{m} \lambda_i \hat{f}_{i\dot{x}}(t) \right) + \sum_{i=0}^{m} \lambda_i \hat{f}_{ix}(t) = 0.$$

D) **Caso no degenerado.** Mostremos que si $\hat{x}(\cdot) \in \operatorname{locextr} \text{Pr}$ será imposible el caso no degenerado. De esta manera el teorema será demostrado.

Elijamos la función $x_j(\cdot) \in C_0^1([t_0, t_1])$ de tal modo que $Ax_j(\cdot) = e_j$, es decir, $\delta\mathcal{J}_i(\hat{x})(\cdot), x_j(\cdot)) = \delta_{ij}$, donde $e_0 = (1, 0, \ldots, 0), \ldots, e_m = (0, \ldots, 0, 1)$ es una base canónica en $\mathbf{R}^{m+1}$. Examinemos la aplicación $\Phi\colon \mathbf{R}^{m+1} \to \mathbf{R}^{m+1}$ que actúa según la fórmula

$$\Phi(\beta) = \Big(\mathcal{J}_0\Big(\hat{x}(\cdot) + \sum_{j=0}^{m} \beta_j x_j(\cdot)\Big),$$

$$\mathcal{J}_1\Big(\hat{x}(\cdot) + \sum_{j=0}^{m} \beta_j x_j(\cdot)\Big), \ldots, \mathcal{J}_m\Big(\hat{x}(\cdot) + \sum_{j=0}^{m} \beta_j x_j(\cdot)\Big)\Big).$$

Es fácil comprobar que la función Φ es continuamente diferenciable en cierto entorno de cero y $\Phi(0) = (\alpha_0, \alpha_1, \ldots, \alpha_m) \overset{\text{def}}{=} \hat{\alpha}\ (\alpha_0 \overset{\text{def}}{=} \mathcal{J}_0(\hat{x}(\cdot)))$. Como la matriz de Jacobi de la aplicación Φ en cero no es degenerada ($\Phi'(0) = (\delta\mathcal{J}_i(\hat{x}(\cdot), x_j(\cdot)))_{i,j_0}^m = 1$ es la matriz unidad), resulta que será aplicable el teorema de función inversa (p. 1.4.4), según el cual existe una aplicación inversa Φ^{-1} de cierto entorno del punto $\hat{\alpha}$, tal que $|\Phi^{-1}(\alpha)| \leqslant K|\alpha - \hat{\alpha}|$ con cierta constante $K > 0$. En particular, para ε bastante pequeño en módulo se hallará un vector $\beta(\varepsilon) = \Phi^{-1}(\alpha_0 + \varepsilon, \alpha_1, \ldots$

$\ldots, \alpha_m)$ tal, para el cual $\Phi(\beta(\varepsilon)) = (\varepsilon + \alpha_0, \alpha_1, \ldots, \alpha_m)$, es decir,

$$\mathcal{J}_0\left(x(\cdot) + \sum_0^m \beta_i(\varepsilon) x_j(\cdot)\right) = \alpha_0 + \varepsilon,$$

$$\mathcal{J}_i\left(\hat{x}(\cdot) + \sum_0^m \beta_j(\varepsilon) x_j(\cdot)\right) = \alpha_i, \quad i = 1, \ldots, m,$$

y en este caso

$|\beta(\varepsilon)| = |\Phi^{-1}(\alpha_0 + \varepsilon, \alpha_1, \ldots, \alpha_m)| \leqslant K|\varepsilon|$.

Resulta que en cualquier entorno de la función $\hat{x}(\cdot)$ (en el espacio $C^1([t_0, t_1])$) existe una función admisible (a saber, $\hat{x}(\cdot) + \sum_{j=0}^m \times$ $\times\, \beta_j(\delta) x_j(\cdot)$ con ε bastante pequeño) cuyo valor de la funcional puede ser mayor o menor que el valor para $\hat{x}(\cdot)$. Hemos llegado a una contradicción con la condición $\hat{x}(\cdot) \in \text{locextr Pr}$. ▷

4.3.3. Ejemplo.

$$\int_0^1 \dot{x}^2\, dt \to \text{extr}; \quad \int_0^1 x\, dt = 0, \quad x(0) = 0, \quad x(1) = 1.$$

Solución: la lagrangiana $L = \lambda_0 \dot{x}^2 - \lambda x$.

Condición necesaria: la ecuación de Euler

$$-\frac{d}{dt} L_{\dot{x}} + L_x = 0 \Leftrightarrow -2\lambda_0 \ddot{x} + \lambda = 0.$$

Si $\lambda_0 = 0$, entonces $\lambda = 0$, es decir, todos los factores de Lagrange son ceros. En este caso no hay extremales admisibles. Sea $\lambda_0 = 1/2$. Entonces $\ddot{x} = \lambda$. Solución general: $x = C_1 t^2 + C_2 t + C_3$. Valiéndonos de las condiciones en los extremos y de las condiciones isoperimétricas, hallamos las constantes incógnitas C_1, C_2 y C_3:

$$x(0) = 0 \Rightarrow C_3 = 0,$$

$$\left.\begin{array}{l} x(1) = 1 \Rightarrow C_1 + C_2 = 1, \\ \displaystyle\int_0^1 x\, dt = 0 \Rightarrow \int_0^1 (C_1 t^2 + C_2 t)\, dt = 0 \Rightarrow \frac{C_1}{3} + \frac{C_2}{2} = 0 \end{array}\right\} \Rightarrow \begin{array}{l} C_1 = 3 \\ C_2 = -2. \end{array}$$

El problema contiene la única extremal admisible, que es $\hat{x} = 3t^2 - 2t$.

Mostremos, a base de verificación directa, que la función $\hat{x}$ proporciona un mínimo absoluto al problema. Tomamos una función $h(\cdot) \in C^1([0, 1])$ tal que $\hat{x}(\cdot) + h(\cdot)$ sea admisible. Para eso debemos tomar una función $h(\cdot)$ para la cual $h(0)=h(1)=0$ y $\int_0^1 h\,dt=0$. En este caso, para la funcional $\mathcal{J}(x(\cdot))\int_0^1 \dot{x}^2\,dt$ tenemos

$$\mathcal{J}(\hat{x}(\cdot)+h(\cdot))-\mathcal{J}(\hat{x}(\cdot))=\int_0^1(\dot{\hat{x}}+\dot{h})^2\,dt-\int_0^1\dot{\hat{x}}^2\,dt=$$

$$=\int_0^1 2\dot{\hat{x}}\dot{h}\,dt+\int_0^1\dot{h}^2\,dt\geqslant 2\int_0^1\dot{\hat{x}}\dot{h}\,dt.$$

Integrando por partes y teniendo en cuenta las condiciones de $h(\cdot)$, obtenemos

$$\mathcal{J}(\hat{x}(\cdot)+h(\cdot))-\mathcal{J}(\hat{x}(\cdot))\geqslant 2\int_0^1\dot{\hat{x}}\,dh=2\dot{\hat{x}}h\Big|_0^1-2\int_0^1\ddot{\hat{x}}h\,dt=$$

$$=-12\int_0^1 h\,dt=0.$$

Es fácil calcular que $S_{\text{mín}}=\int_0^1\dot{\hat{x}}^2\,dt=\int_0^1(6t-2)^2\,dt=4.$

Resultado. La función $\hat{x}=3t^2-2t$ proporciona al problema un mínimo absoluto. El valor de tal problema constituye $S_{\text{mín}}=4$. Es obvio que $S_{\text{máx}}=+\infty$.

§ 5. Problema de Lagrange

Este párrafo está dedicado a la deducción de las condiciones necesarias de extremo débil para una sola clase general de problemas del cálculo clásico de variaciones, es decir, a los problemas de Lagrange. A esta clase pertenecen todos los problemas estudiados en el párrafo anterior. La deducción de las condiciones necesarias se basa en los resultados del análisis de dimensión infinita o, mejor dicho, en el principio lagrangiano para los problemas suaves con acotaciones del tipo de igualdades y desigualdades.

5.1. Principio lagrangiano para el problema de Lagrange

5.1.1. Planteamiento del problema. Sea $X=C^1(\Delta,\ \mathbf{R}^n)\times\mathbf{R}^2$, donde Δ es un segmento finito dado. La condición $\xi=(x(\cdot),\ t_0,\ t_1)\in X$ significa que $x(\cdot)\in C^1(\Delta,\ \mathbf{R}^n)$, $t_0,\ t_1\in\Delta$. Seguidamente, a ciertas coordenadas del vector función $x(\cdot)=(x_1(\cdot),\ \ldots,\ x_n(\cdot))$ (para precisar consideraremos que son las primeras k coordenadas) estará aplicada la relación diferencial $\dot{x}_i = \varphi_i(t, x(t))$, $i = 1, \ldots, k$. Denotemos a continuación $x(\cdot) = (x_\alpha(\cdot), x_\beta(\cdot))$, donde $x_\alpha(\cdot) = (x_1(\cdot), \ldots x_k(\cdot))$, $x_\beta(\cdot) = (x_{k+1}(\cdot), \ldots, x_n(\cdot))$. Si no existe relación diferencial, entonces $k = 0$ y $x(\cdot) = x_\beta(\cdot)$. Llámase problema de Lagrange el siguiente problema extremal en el espacio X:

$$\mathscr{B}_0(\xi) \to \inf: \mathscr{B}_i(\xi) \leqslant 0, \ i = 1, \ldots, m' \tag{1}$$

$$\mathscr{B}_i(\xi) = 0, \ i = m' + 1, \ldots, m, \tag{Pr}$$

$$\Phi(\xi) = \dot{x}_\alpha(t) - \varphi(t, x(t)) = 0 \ \forall \ t \in \Delta, \tag{2}$$

donde

$$\mathscr{B}_i(\xi) = \int_{t_0}^{t_1} f_i(t, x(t), \dot{x}(t))\, dt - \psi_i(t_0, x(t_0), t_1, x(t_1)),$$

$$i = 0, 1, \ldots, m, \ t_0, t_1 \in \operatorname{int} \Delta, \quad t_0 < t_1.$$

Se consideran casos parciales (Pr) los problemas en los que uno de los extremos t_0 o t_1 es móvil, y el otro está fijo o ambos extremos $[t_0$ y $t_1]$ del segmento permanecen fijos.

El elemento ξ para el cual se cumplen las condiciones y acotaciones (1)—(2) del tipo de igualdades y desigualdades, se llama *admisible*. El conjunto de los demás elementos admisibles constituyen un subconjunto en el espacio X.

Supongamos que el elemento admisible $\hat{\xi} = (\hat{x}(\cdot), \hat{t}_0, \hat{t}_1)$ proporciona *un mínimo local débil* al problema de Lagrange (Pr), y escribimos $\hat{\xi} \in$ locmín Pr si existe $\delta > 0$ tal que para cualquier elemento admisible $\xi = (x(\cdot), t_0, t_1)$ que satisfaga la condición $\| \xi - \hat{\xi} \|_X < \delta$ $(\Leftrightarrow \| x(\cdot) - \hat{x}(\cdot) \|_{C^1(\Delta, \mathbf{R}^n)} < \delta, \ | t_0 - \hat{t}_0 < \delta, \ | t_1 - \hat{t}_1 | < \delta)$ se cumple la desigualdad $\mathscr{B}_0(\xi) \geqslant \mathscr{B}_0(\hat{\xi})$.

En vista de que en la función $f_i(t, x, \dot{x})$, en vez de $\dot{x}_\alpha$ se puede poner la expresión $\varphi(t, x)$ tomada de (2) e igual a la primera, en lo sucesivo consideraremos que $f_i = f_i(t, x, \dot{x}_\beta)$.

El planteamiento del problema de Lagrange de la forma indicada, así como la formulación posterior de las condiciones necesarias, permiten, en una serie de casos, simplificar, en comparación con el problema lagrangiano en ATF y AGT, la escritura de las condiciones precisas (la cantidad de ecuaciones diferenciales puede ser reducida en 1).

5.1.2. Condiciones de extremo necesarias.

Teorema de Euler — Lagrange. *Supongamos que el elemento* $\hat{\xi} = (\hat{x}(\cdot), \hat{t}_0, \hat{t}_1$ *proporciona un mínimo local débil al problema de Lagrange* (Pr). *En este caso son continuas, en cierto entorno del con-*

junto $\{(t, \hat{x}(t), \hat{x}_\beta(t)) \mid t \in \Delta\}$, *las funciones* f_i, $i = 0, 1, \ldots, m$ *y sus derivadas parciales en* $\dot{x}$ *y* x, *así como* φ *y su derivada parcial en* x *en cierto entorno del conjunto* $\{(t, \hat{x}(t)) \mid t \in \Delta\}$, *y son continuamente diferenciables las funciones* ψ_i, $i = 0$ $1, \ldots, m$ *en el entorno del punto* $(\hat{t}_0, \hat{x}(\hat{t}\ ), \hat{t}_1, \hat{x}(\hat{t}_1))$ *(condición de suavidad).*

Entonces se hallarán los factores de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots$ $\ldots, \lambda_m) \in \mathbf{R}^{m+1}$ *y la función vectorial* $p(\cdot) \in C^1$ $\Delta, \mathbf{R}^k)$ *que no son iguales a cero simultáneamente, pero que son tales que para la función lagrangiana*

$$\mathcal{L} = \int_{t_0}^{t_1} [f(t, x, \dot{x}) + \langle p(t), \dot{x}_\alpha - \varphi(t, x(t))\rangle] dt + l(t_0, x(t_0), t_1, x(t_1)),$$

donde

$$f(t, x, \dot{x}) = \sum_{i=0}^{m} \lambda_i f_i(t, x, \dot{x}_\beta),$$

$$l = \sum_{i=0}^{m} \lambda_i \psi_i(t_0, x(t_0), t_1, x(t_1)),$$

se cumplen las condiciones:

a) *de estacionaridad en* x, *o sea, la ecuación de Euler para la lagrangiana*

$$L(t, x, \dot{x}) = f(t, x, \dot{x}_\beta) + \langle p(t), \dot{x}_\alpha - \varphi(t, x(t))\rangle - \frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) =$$

$$= 0 \Leftrightarrow \begin{cases} -\dot{p}(t) - p(t)\hat{\varphi}_{x_\alpha}(t) + \hat{f}_{x_\alpha}(t) = 0, \\ -\dfrac{d}{dt}\hat{f}_{\dot{x}_\beta}(t) + \hat{f}_{x_\beta}(t) - p(t)\hat{\varphi}_{x_\beta}(t) = 0; \end{cases}$$

b) *de transversalidad en* x:

$$\hat{L}_{\dot{x}}(\hat{t}_i) = (-1)^i \hat{l}_{x(t_i)} \Rightarrow \begin{cases} p(\hat{t}_i) = (-1)^i \hat{l}_{x_\alpha(t_i)}; \\ \hat{f}_{\dot{x}_\beta}(\hat{t}_i) = (-1)^k \hat{l}_{x_\beta(t_i)}, \\ i = 0, 1; \end{cases}$$

c) *de estacionaridad en* t_0, t_1 *(sólo se escribe para los extremos móviles del segmento de integración):*

$$\hat{\mathcal{L}}_{t_0} = 0 \Leftrightarrow -\hat{f}(\hat{t}_0) + l_{t_0} + l_{x(t_0)}\dot{\hat{x}}(\hat{t}_0) = 0,$$

$$\hat{\mathcal{L}}_{t_1} = 0 \Leftrightarrow \hat{f}(\hat{t}_1) + \hat{l}_{t_1} + \hat{l}_{x(t_1)}\dot{\hat{x}}(t_1) = 0;$$

d) *de no rigidez complementaria:*

$$\lambda_i \mathcal{B}_i(\hat{\xi}) = 0, \mid \quad i = 1, \ldots, m';$$

e) *de no negatividad*;

$$\lambda_i \geqslant 0, \quad i = 0, 1, \ldots, m'.$$

◁ Examinemos el problema (Pr) como un problema extremal con acotaciones del tipo de igualdades y desigualdades (p. 2.4). En nuestro caso el espacio $X = C^1(\Delta, \mathbf{R}^n) \times \mathbf{R}^2$ y las funcionales $\mathscr{B}_i$, $i = 1, \ldots, m'$ proporcionan acotaciones del tipo de desigualdades, mientras que la acotación del tipo de igualdad es proporcionada por la aplicación $F(\xi) = (\Phi(\xi), \mathscr{B}_{m'+1}(\xi), \ldots, \mathscr{B}_m(\xi))$ que actúa en el espacio $Y = C(\Delta, \mathbf{R}^k) \times \mathbf{R}^{m-m'}$. A su vez, $\hat{\xi}$ proporciona un mínimo local al espacio X.

Mostremos que en nuestro problema (Pr) se cumplen todas las condiciones del teorema del p. 2.4 (su carácter banachiano, suavidad y regularidad debilitada), y después, según el principio de Lagrange para los problemas con igualdades y desigualdades, escribamos las condiciones de mínimo necesarias y deduzcamos de ellas las condiciones a) — e) de nuestro teorema.

El carácter banachiano de los espacios X e Y se deduce del hecho de que los espacios $C^1(\Delta, \mathbf{R}^n)$ y $C(\Delta, \mathbf{R}^k)$ son banachianos (p. 1.1.2), así como de que el producto de los espacios banachianos también es banachiano (p. 1.1.3).

La suavidad de las aplicaciones $\mathscr{B}_i$ y F se cumple debido a que de la diferenciabilidad continua en el entorno del punto $\hat{\xi}$ se deduce una diferenciabilidad estricta en el punto (p. 1.4.2).

La regularidad debilitada, o sea, el carácter cerrado del conjunto $F'(\hat{\xi})X$ se desprende del lema sobre el carácter cerrado de la imagen (p. 7.2), donde $A = \Phi'(\hat{\xi})$ $(A: X \to Y)$, $B = (\mathscr{B}'_{m'+1}(\hat{\xi}), \ldots$ $\ldots, \mathscr{B}'_k(\hat{\xi}))$ $(B: X \to \mathbf{R}^{m-m'})$. El conjunto $B\operatorname{Ker} A$ está cerrado ya que es el subespacio de un espacio de dimensión finita. El carácter cerrado de la imagen de la aplicación A tiene lugar en virtud de que la imagen $\Phi'(\hat{\xi})C^1(\Delta, \mathbf{R}^n)$ simplemente coincide con $C(\Delta, \mathbf{R}^k)$. En efecto, tomando arbitrariamente $y(\cdot) \in C(\Delta, \mathbf{R}^k)$ (y el vector arbitrario $\gamma \in \mathbf{R}^k$) siempre podemos hallar la función $(h(\cdot) \in C^1(\Delta, \mathbf{R}^k)$ y, asimismo, la función $(h(\cdot), 0)$ $(h_\beta(\cdot) = 0)$ del espacio $C^1(\Delta, \mathbf{R}^n)$, para la cual

$$\Phi'(\hat{\xi})[(h(\cdot), 0] = \dot{h}(t) - \hat{\varphi}_{x_\alpha}(t)h(t) = y(t), \quad h(\hat{t}_0) = \gamma. \tag{1}$$

La solución del último sistema de k ecuaciones lineales diferenciales con factores continuos existe, es única y está definida en todo el segmento Δ en virtud del teorema de existencia y unicidad de la solución del teorema de Cauchy para un sistema lineal heterogéneo [ATF, pág. 191].

Por consiguiente, todas las condiciones del teorema (p. 2.4) se cumplen. Conforme a ese teorema se hallarán los factores de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbf{R}^{m+1}$, $y^* \in Y^*$ que no todos son iguales a cero, pero que son tales que para la función lagrangiana del problema (Pr)

$$\widetilde{\mathscr{L}}(\xi, y^*, \lambda) = \widetilde{\mathscr{L}}(x(\cdot), t_0, t_1, y^*, \lambda) =$$

$$= \int_{t_0}^{t_1} f(t, x, \dot{x}_\beta)\, dt + \langle y^*, \dot{x}_\alpha - \varphi(t, x(t))\rangle + l(t_0, x(t_0), t_1, x(t_1))$$

se cumplen las condiciones de estacionaridad

$$\hat{\widetilde{\mathscr{L}}}(\hat{\xi}) = 0 \Leftrightarrow \hat{\widetilde{\mathscr{L}}}_x = 0 \,(\Leftrightarrow \hat{\widetilde{\mathscr{L}}}_{x_\alpha} = 0,\ \hat{\widetilde{\mathscr{L}}}_{x_\beta} = 0),$$

$$\hat{\widetilde{\mathscr{L}}}_{t_0} = 0, \quad \hat{\widetilde{\mathscr{L}}}_{t_1} = 0,$$

de no rigidez complementaria y de no negatividad.

Mostremos que de la igualdad $\hat{\mathscr{L}}_{x_\alpha} = 0$ se deduce la existencia de la función $p(\cdot) \in C^1(\Delta, \mathbf{R}^h)$ para la cual se cumplen las condiciones a), b) del teorema y, además, $\widetilde{\mathscr{L}} = \mathscr{L}$. En este caso el teorema será demostrado.

Descifremos la condición de estacionaridad de $\widetilde{\mathscr{L}}$ en x_α:

$$0 = \hat{\widetilde{\mathscr{L}}}_x[h(\cdot)] = \int_{\hat{t}_0}^{\hat{t}_1} \hat{f}_{x_\alpha} h\, dt + \langle y^*, \dot{h}\rangle - \varphi_{x_\alpha} h\rangle +$$

$$+ \hat{l}_{x_\alpha}(t_0)\, h(\hat{t}_0) + \hat{l}_{x_\alpha}(t_1)\, h(\hat{t}_1)$$

$$\forall h \in C^1(\Delta, \mathbf{R}^h).$$

De aquí, en virtud de la expresión 1,

$$\langle y^*, y(\cdot)\rangle = -\int_{\hat{t}_0}^{\hat{t}_1} \hat{f}_{x_\alpha} h\, dt - \hat{l}_{x_\alpha}(t_0)\,\gamma - l_{x_\alpha}(t_1)\, h(\hat{t}_1)\ \forall y \in C^1(\Delta, \mathbf{R}^h),$$

$$\forall \gamma \in \mathbf{R}^h. \tag{2}$$

Determinemos la función p de las condiciones

$$\dot{p}(t) + p(t)\,\hat{\varphi}_{x_\alpha}(t) = f_{x_\alpha}(t), \quad p(\hat{t}_1) = -\hat{l}_{x_\alpha}(t_1). \tag{3}$$

En virtud del teorema de existencia y unicidad de la solución del problema de Cauchy para un sistema lineal heterogéneo [ATF, pág. 191], la función $p(\cdot) \in C^1(\Delta, \mathbf{R}^h)$ será determinada unívocamente a base de nuestras condiciones. Por lo tanto, debido a las condiciones (1) y (3), tenemos

$$\int_{\hat{t}_0}^{\hat{t}_1} \frac{d}{dt}(ph)\,dt = p(\hat{t}_1)h(\hat{t}_1) - p(\hat{t}_0)h(\hat{t}_0) = \int_{\hat{t}_0}^{\hat{t}_1} (\dot{p}h + p\dot{h})\,dt =$$

$$= \int_{\hat{t}_0}^{\hat{t}_1} (\hat{f}_{x_\alpha}h - p\hat{\varphi}_{x_\alpha}h + p\hat{\varphi}_{x_\alpha}h + py)\,dt = \int_{\hat{t}_0}^{\hat{t}_1} (\hat{f}_{x_\alpha}h + py)\,dt.$$

Hallando $\int_{\hat{t}_0}^{\hat{t}_1} \hat{f}_{x_\alpha}h\,dt$ de la última relación e introduciendo la expresión obtenida en la identidad (2), obtenemos

$$\langle y^*, y\rangle = \int_{\hat{t}_0}^{\hat{t}_1} py\,dt - (\hat{l}_{x_\alpha}(t_0) - p(\hat{t}_0))\gamma$$

$$\forall y \in C(\Delta, \mathbf{R}^k), \quad \forall \gamma \in \mathbf{R}^k,$$

de donde se deduce que $\langle y^*, y\rangle = \int_{\hat{t}_0}^{\hat{t}_1} p(t)\,y(t)\,dt$, $p(\hat{t}_0) = \hat{l}_{x_\alpha}(t_0)$.

Por consiguiente, $\tilde{\mathscr{L}} = \mathscr{L}$ y resultan demostradas las ecuaciones de Euler y las condiciones de transversalidad de las variables x_α.

La ecuación euleriana y la condición de transversalidad para las variables x_β se deducen de la estacionaridad de la función lagrangiana $\mathscr{L}$ en x_β. El correspondiente resultado fue obtenido al deducir las condiciones necesarias en el problema de Bolz (p. 4.1.2). ▷

5.1.3. Ejemplo. $\int_0^1 \ddot{x}^2\,dt \to \text{extr}$; $x(0) = \dot{x}(0) = 0$, $x(1) = 1$.

Solución. Reduzcamos el problema a la forma de problema de Lagrange (p. 5.1.1), sustituyendo las variables $x_1 = x$, $x_2 = \dot{x}$:

$$\int_0^1 \dot{x}_2^2\,dt \to \text{extr}; \quad \dot{x}_1 = x_2, \quad x_1(0) = x_2(0) = 0,$$

$$x_1(1) = 1.$$

La función de Lagrange:

$$\mathscr{L} = \int_0^1 (\lambda_0\dot{x}_2^2 + p(\dot{x}_1 - x_2))\,dt + \lambda_1 x_1(0) + \lambda_2 x_2(0) + \lambda_3(x_1(1) - 1).$$

Condiciones necesarias:

a) sistema de ecuaciones de Euler para la lagrangiana $L = \lambda_0 \dot{x}_0^2 + p(x_1 - x_2)$:

$$-\frac{d}{dt} L_{\dot{x}_i} + L_{x_i} = 0, \quad i = 1,\ 2 \Leftrightarrow -\dot{p} = 0, \quad -2\lambda_0 \ddot{x}_2 - p = 0;$$

b) transversalidad en x para el terminante $l = \lambda_1 x_1(0) + \lambda_2 x_2(0) + \lambda_3 (x_1(1) - 1)$:

$$L_{\dot{x}_i}(0) = l_{x_i(0)}, \ i = 1,\ 2, \ p(0) = \lambda_1, \ p(1) = -\lambda_3,$$

$$L_{\dot{x}_i} = -l_{x_i(1)}, \quad i = 1,\ 2, \ 2\lambda_0 \dot{x}_2(0) = \lambda_2,$$

$$2\lambda_0 \dot{x}_2(1) = 0.$$

En vista de que los extremos del segmento de integración son fijos, no es necesario escribir la condición de estacionaridad en t_0, t_1. Además, también faltan las condiciones de no rigidez complementaria y de no negatividad, puesto que no hay acotaciones del tipo de desigualdades.

Si $\lambda_0 = 0$, de b) y a) se deduce que $\lambda_2 = 0$, $p \equiv 0$. Además, de b) se deduce que $\lambda_1 = \lambda_3 = 0$, es decir, que todos los factores de Lagrange son ceros. Así, para $\lambda_0 = 0$ no hay extremales admisibles. Sea $\lambda_0 \neq 0$. Del sistema de ecuaciones eulerianas se deduce que $\dddot{x}_2 = 0$, lo cual equivale a la ecuación $\overset{....}{x} = 0$, de donde $x(t) = C_1 t^3 + C_2 t^2 + C_3 t + C_4$. Las constantes incógnitas $C_1 - C_4$ se determinan de las condiciones en los extremos y de la condición de transversalidad $\ddot{x}(1) = 0$. La única extremal admisible es $\hat{x}(t) = (1/2)(3t^2 - t^3)$.

Mostremos, mediante verificación directa, que $\hat{x} \in$ absmín Pr. Tomemos una función $h(\cdot) \in C^2([0, 1])$ tal que $\hat{x}(\cdot) + h(\cdot)$ sea admisible en el problema. Para eso es necesario tomar la función $h(\cdot)$ con las siguientes condiciones en los extremos: $h(0) = \dot{h}(0) = h(1) = 0$. Para la funcional $\mathscr{B}(x(\cdot)) = \int_0^1 \ddot{x}^2\, dt$ tenemos

$$\mathscr{B}(\hat{x}(\cdot) + h(\cdot)) - \mathscr{B}(\hat{x}(\cdot)) = \int_0^1 (\ddot{\hat{x}} + \ddot{h})^2\, dt - \int_0^1 \ddot{\hat{x}}^2\, dt =$$

$$= 2\int_0^1 \ddot{\hat{x}}\ddot{h}\, dt + \int_0^1 \ddot{h}^2\, dt \geqslant 2\int_0^1 \ddot{\hat{x}}\ddot{h}\, dt.$$

Después, integrando dos veces por partes y teniendo en cuenta las condiciones en los extremos de las funciones $\hat{x}$ y h y la ecuación

$\ddddot{\hat{x}} = 0$, obtenemos

$$\mathcal{B}(\hat{x}(\cdot)+h(\cdot))-B(\hat{x}(\cdot))\geqslant 2\int_0^1 \ddot{\hat{x}}\,d\dot{h}=$$

$$=2\ddot{\hat{x}}\dot{h}\Big|_0^1-2\int_0^1 \dddot{\hat{x}}\,\dot{h}\,dt=-2\int_0^1 \dddot{\hat{x}}\,dh=-2\dddot{\hat{x}}\ \Big|\ \ddddot{\hat{x}}\,h\,dt=0.$$

Así pues, $\mathcal{B}(\hat{x}(\cdot)+h(\cdot))\geqslant\mathcal{B}(\hat{x}(\cdot))$ y, por consiguiente, $\hat{x}(\cdot)\in$ absmín Pr, $S_{\text{mín}}=\mathcal{B}(\hat{x}(\cdot))=3$. Está claro que $S_{\text{máx}}=$ $=+\infty$. Efectivamente, tomamos la sucesión $x_n(t)=\hat{x}(t)+nh(t)$, donde $h(\cdot)$ es cierta función de $C_0^2([0,\ 1])$, tal que $\ddot{h}\not\equiv 0$, por ejemplo, $h(t)=t^2(t-1)^2$. Entonces $\mathcal{B}(x_n(\cdot))\to+\infty$ para $n\to\infty$.

5.2. Problema de extremos móviles

5.2.1. Planteamiento del problema. Llámase *problema de extremos móviles* el siguiente problema en el espacio $C^1(\Delta)\times\mathbf{R}^2$:

$$\mathcal{J}(x(\cdot),\ t_0,\ t_1)=\int_{t_0}^{t_1} L(t,\ x(t),\ \dot{x}(t))\,dt+$$
$$+\psi_0(t_0,\ x(t_0),\ t_1,\ x(t_1))\to\text{extr};\tag{Pr}$$
$$\psi_i(t_0,\ x(t_0),\ t_1,\ x(t_1))=0,\quad i=1,\ \ldots,\ m,$$

donde Δ es un segmento finito dado, $t_0,\ t_1\in\Delta$, $t_0<t_1$.

En el caso parcial, Pr es un problema donde uno de los extremos, es decir, t_0 o t_1, es móvil, y el otro es fijo.

El conjunto triple $(x(\cdot),\ t_0,\ t_1)$ se llama admisible en (Pr) si $x(\cdot)\in C^1(\Delta)$, $t_0,\ t_1\in\operatorname{int}\Delta$, $t_0<t_1$ y si se cumplen las condiciones (1) en los extremos.

Decimos que el conjunto triple admisible $(\hat{x}(\cdot),\ \hat{t}_0,\ \hat{t}_1)$ proporciona un mínimo (máximo) local débil al problema (Pr) (en el espacio $C^1(\Delta)\times\mathbf{R}^2$) si existe $\delta>0$ tal que con cualquier otro conjunto triple admisible $(x(\cdot),\ t_0,\ t_1)$, para el cual $|t_0-\hat{t}_0|\leqslant\delta$, $|\hat{t}_1-$ $-t_1|\leqslant\delta$, $\|x(\cdot)-\hat{x}(\cdot)\|_{C^1(\Delta)}\leqslant\delta$, se cumple la desigualdad

$$\mathcal{J}(x(\cdot),\ t_0,\ t_1)\geqslant\mathcal{J}(\hat{x}(\cdot),\ \hat{t}_0,\ \hat{t}_1)$$
$$(\mathcal{J}(x(\cdot),\ t_0,\ t_1)\leqslant\mathcal{J}(\hat{x}(\cdot),\ \hat{t}_0,\ \hat{t}_1)).$$

En este caso escribimos $(\hat{x}(\cdot),\ \hat{t}_0,\ \hat{t}_1)\in$ locmín Pr (locmáx Pr).

5.2.2. Condiciones de extremo necesarias. Teorema. *Supongamos que el conjunto triple* $(\hat{x}(\cdot), \hat{t}_0, \hat{t}_1)$ *proporciona un extremo local débil al problema de extremos móviles* (Pr). *Además, el integrante L y sus derivadas parciales en x y* $\dot{x}$ *son continuos en cierto entorno del conjunto* $\{(t, \hat{x}(t), \dot{\hat{x}}(t)) \mid t \in \Delta\}$, *mientras que las funciones* ψ_i, $i = 0, 1, \ldots, m$ *son continuamente diferenciables en el entorno del punto* $(\hat{t}_0, \hat{x}(\hat{t}_0), \hat{t}_1, \hat{x}(\hat{t}_1))$.

Entonces se hallará un factor no nulo de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbf{R}^{m+1}$ *tal que para la función lagrangiana*

$$\mathcal{L} = \int_{t_0}^{t_1} \lambda_0 L(t, x, \dot{x})\, dt + l(t_0, x(t_0), t_1, x(t_1)),$$

donde $l = \sum_{i=0}^{m} \lambda_i \psi_i (t_0, x(t_0), t_1, x(t_1))$, *se cumplen las condiciones:*

a) *de estacionaridad en x, es decir, la ecuación de Euler para el integrante* $\lambda_0 L(t, x, \dot{x})$:

$$-\frac{d}{dt} \lambda_0 \hat{L}_{\dot{x}}(t) + \lambda_0 \hat{L}_x(t) = 0;$$

b) *de transversalidad en x:*

$$\lambda_0 \hat{L}_{\dot{x}}(\hat{t}_0) = \hat{l}_x, \quad \lambda_0 \hat{L}_{\dot{x}}(\hat{t}_1) = -\hat{l}_{x(t_1)};$$

c) *de estacionaridad en* t_0, t_1 *(sólo se escribe para los extremos móviles):*

$$\hat{\mathcal{L}}t_0 = 0 \Leftrightarrow -\lambda_0 \hat{L}(\hat{t}_0) + \hat{l}_{t_0} + \hat{l}_{x(t_0)} \dot{\hat{x}}(\hat{t}_0) = 0,$$

$$\hat{\mathcal{L}}t_1 = 0 \Leftrightarrow \lambda_0 \hat{L}(\hat{t}_1) + \hat{l}_{t_1} + \hat{l}_{x(t_1)} \dot{\hat{x}}(\hat{t}_1) = 0.$$

◁ Este teorema se deduce directamente del principio lagrangiano para el problema de Lagrange donde falta la relación diferencial y las acotaciones del tipo de desigualdades. ▷

5.2.3. Ejemplo. $\mathcal{J}(x(\cdot), T) = \int_0^T (\dot{x}^2 - x + 1)\, dt \to \text{extr}; \ x(0) = 0.$

Solución. Función lagrangiana:

$$\mathcal{L} = \int_0^T \lambda_0 (\dot{x}^2 - x + 1)\, dt + \lambda x(0).$$

Condiciones necesarias:

a) ecuación de Euler para el integrante $L = \lambda_0 (\dot{x}^2 - x + 1)$:

$$-\frac{d}{dt} L_{\dot{x}} + L_x = 0 \Leftrightarrow \lambda_0 (2\ddot{x} + 1) = 0;$$

b) condiciones de transversalidad en x para el terminante $l = \lambda x(0)$:

$$L_{\dot{x}}(0) = l_{x(0)},\quad L_{\dot{x}}(T) = -l_{x(T)} \Leftrightarrow 2\lambda_0 \dot{x}(0) = \lambda,$$
$$2\lambda_0 \dot{x}(\hat{T}) = 0;$$

c) condición de estacionaridad en T (sólo se escribe para el extremo móvil):

$$\mathscr{L}_T(\hat{T}) = 0 \Leftrightarrow \lambda_0 (\dot{x}^2(\hat{T}) - x(\hat{T}) + 1) = 0.$$

Si $\lambda_0 = 0$, de b) se deduce que $\lambda = 0$, o sea, que todos los factores lagrangianos son ceros. Sea $\lambda_0 = 1$. Entonces, de a) se deduce que $\ddot{x} = -1/2$. La solución general de dicha ecuación diferencial resulta $x = (-t^2/4) + C_1 t + C_2$. En vista de que $x(0) = 0$, obtenemos $C_2 = 0$. Mediante las ecuaciones $\dot{x}(\hat{T}) = 0$ y $\dot{x}^2(\hat{T}) - \hat{x}(\hat{T}) + 1 = 0$ determinamos las incógnitas C_1 y $\hat{T}$. Resolviendo ese sistema de ecuaciones hallamos que $\hat{T} = 2$, $C_1 = 1$.

El problema tiene una sola extremal admisible $\hat{x} = -t^2/4 + t$ expresada en el segmento $[0, 2]$.

Mostremos que $(\hat{x}(\cdot), \hat{T}) \notin$ locextr. Efectivamente, para la función $\hat{x}(t) = -t^2/4 + t$,

$$\mathcal{J}(\hat{x}(\cdot), T) = \int_0^T (\dot{\hat{x}}^2 - \hat{x} + 1)\, dt =$$
$$= \int_0^T \left(\left(-\frac{t}{2} + 1 \right)^2 - \left(\frac{t^2}{4} + t \right) + 1 \right) dt = \int_0^T \frac{(t-2)^2}{2}\, dt.$$

Cuando T son próximos a $\hat{T}$, los valores de la funcional $\mathcal{J}(\hat{x}(\cdot), T)$ pueden ser menores o mayores de $\mathcal{J}(\hat{x}(\cdot), \hat{T})$.

Tomemos la sucesión de pares $x_n(t) = t$, $T_n = n$; entonces $\mathcal{J}(x_n(\cdot), T_n) \to -\infty$ con $n \to \infty$. Por lo tanto, $S_{\text{mín}} = -\infty$. Evidentemente, $S_{\text{máx}} = +\infty$.

5.3. Problemas de derivadas superiores

5.3.1. Planteamiento del problema. Llámase *problema de derivadas superiores* (con extremos fijos) en el cálculo clásico de variaciones, el siguiente problema en el espacio $C^n([t_0, t_1])$:

$$\mathcal{J}(x(\cdot)) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t), \ldots, x^{(n)}(t)\, dt \to \text{extr}; \qquad \text{(Pr)}$$

$$x^{(k)}(t_j) = x_j^k, \quad k = 0, 1, \ldots, n-1, \quad j = 0, 1. \qquad (1)$$

Aquí L: $\mathbf{R}^{n+2} \to \mathbf{R}$ es una función de $n+2$ variables. La función L se llama *integrante*. Las funciones $x(\cdot) \in C^n([t_0, t_1])$ que satisfacen las condiciones (1) en los extremos del segmento $[t_0, t_1]$ se llaman *admisibles*.

Decimos que la función admisible $\hat{x}(\cdot)$ proporciona al problema (Pr) *un mínimo (máximo) local* en el espacio $C^n([t_0, t_1])$, si existe $\delta > 0$ tal que con cualquier función admisible $x(\cdot)$, para la cual $\| x(\cdot) - \hat{x}(\cdot) \|_{C^n([t_0, t_1])} < \delta$, se cumple la desigualdad

$$\mathcal{J}(x(\cdot)) \geqslant \mathcal{J}(\hat{x}(\cdot)) \quad (\mathcal{J}(x(\cdot)) \leqslant \mathcal{J}(\hat{x}(\cdot))),$$

y en este caso escribimos $\hat{x}(\cdot) \in \text{locmín Pr}$ (locmáx Pr).

5.3.2. Ecuación de Euler — Poisson. Teorema. *Supongamos que la función $x(\cdot)$ proporciona un extremo local al problema de derivadas superiores* (Pr). *Además, el integrante L es tal que $\hat{L}_{x(k)}(\cdot) \in C^k([t_0, t_1])$, $k = 0, 1, \ldots, n$ (condición de suavidad). Entonces en la extremal $\hat{x}(\cdot)$ se cumple la ecuación de Euler — Poisson:*

$$\sum_{k=0}^{n} (-1)^k \left(\frac{d}{dt}\right)^k \hat{L}_{x(k)}(t) = 0.$$

◁ Reduzcamos el problema de derivadas superiores al problema de Lagrange, sustituyendo las variables $x_k = x^{(k-1)}$, $k = 1, \ldots, n$,

$$\int_{t_0}^{t_1} L(t, x_1, x_2, \ldots, x_n, \dot{x}_n)\, dt \to \text{extr};$$

$$\dot{x}_k = x_{k+1}, \; k = 1, \ldots, n, \; x_k(t_j) = x_j^{k-1}, \; j = 0, 1,$$

$$k = 1, \ldots, n. \qquad \text{(Pr}')$$

Aquí la variable fásica es la función vectorial $(x_1(\cdot), \ldots, x_n(\cdot))$. En vista de que la función $x(\cdot)$ proporciona un extremo local al problema de derivadas superiores (Pr), la función vectorial $(\hat{x}(\cdot), \ldots, \hat{x}_n(\cdot))$ proporcionará un extremo local débil al problema de Lagrange (Pr$'$). Con arreglo al teorema de Euler — Lagrange escribimos las condiciones necesarias de estacionaridad para la lagrangiana $\tilde{L} = \lambda_0 L(t, x_1, \ldots, x_n, \dot{x}_n) + \sum_{k=1}^{n-1} p_k (\dot{x}_k - x_{k+1})$. No escribiremos la parte terminal de la función lagrangiana, ni tampoco las demás condiciones necesarias de extremo que no desempeñan un papel importante en el problema de extremos fijos y de segmento de integración fijo. A continuación se ofrece el sistema de ecuaciones de Euler:

$$-\dot{p}_1+\lambda_0 \frac{\partial \hat{L}}{\partial x_1}=0,$$

$$-\frac{d}{dt}\hat{L}_{\dot{x}_k}(t)+\hat{L}_{x_k}(t)=0 \Leftrightarrow -\dot{p}_k+\lambda_0 \frac{\partial \hat{L}}{\partial x_k}-p_{k-1}=0,$$

$$k=1, \ldots, n, \; k=2, \ldots, n-1,$$

$$-\frac{d}{dt}\lambda_0 \frac{\partial \hat{L}}{\partial x_n}+\lambda_0 \frac{\partial \hat{L}}{\partial x_n}-P_{n-1}=0.$$

Si $\lambda_0 = 0$, del sistema de ecuaciones de Euler se deduce que $p_{n-1} = \ldots = p_1 = 0$. Todos los factores de Lagrange son ceros. Sea $\lambda_0 \neq 0$. Supongamos que $\lambda_0 = 1$. Expresemos P_{n-1} de la última ecuación e introduzcámoslo en la penúltima ecuación. Después de realizar este procedimiento con arreglo a $p_{n-2}, \ldots, p_1$ llegaremos, al fin y al cabo, a la ecuación de Euler — Poisson.

5.3.3. Ejemplo.

$$\int_0^1 \ddot{x}^2\,dt \to \text{extr}; \quad x(0)=\dot{x}(0)=x(1)=0, \; \dot{x}(1)=1.$$

Solución. Integrante: $L = \ddot{x}^2$.

La condición necesaria es la ecuación de Euler — Poisson

$$\frac{d^2}{dt^2}L_{\ddot{x}}-\frac{d}{dt}L_{\dot{x}}+L_x=0 \Leftrightarrow x^{(4)}(t)=0.$$

Solución general de la ecuación de Euler — Poisson:

$$x(t) = C_1 t^3 + C_2 t^2 + C_3 t + C_4.$$

Las constantes incógnitas C_1, C_2, C_3, C_4 se determinan de las condiciones de contorno

$$x(0) = 0 \Rightarrow C_4 = 0,$$

$$\dot{x}(0) = 0 \Rightarrow C_3 = 0,$$

$$\left.\begin{aligned} x(1)=0 &\Rightarrow C_1+C_2=0,\\ \dot{x}(1)=1 &\Rightarrow 3C_1+2C_2=1 \end{aligned}\right\} \Rightarrow \begin{aligned} C_1&=1,\\ C_2&=-1. \end{aligned}$$

Eso significa que el problema sólo tiene una extremal admisible: $\hat{x} = t^3 - t^2$.

Mostremos que dicha extremal proporciona un mínimo absoluto al problema. Efectivamente, si $h(\cdot) \in C_0^2[0, 1]$, entonces

$$\mathcal{J}(\hat{x}(\cdot)+h(\cdot))-\mathcal{J}(\hat{x}(\cdot))=\int_0^1 (\ddot{\hat{x}}+\ddot{h})^2\,dt-\int_0^1 \ddot{\hat{x}}\,dt=$$

$$=2\int_0^1 \ddot{\hat{x}}\ddot{h}\,dt+\int_0^1 \ddot{h}^2\,dt,$$

Integrando dos veces por partes y estimando que $h(0)=h(1)=$ $=\dot h(0)=\dot h(1)=0$, obtenemos

$$\int_0^1 \ddot{\hat x}\,\ddot h\,dt=\int_0^1 \ddot{\hat x}\,d\dot h=\ddot{\hat x}\dot h\Big|_0^1-\int_0^1 \dddot{\hat x}\dot h\,dt=\int_0^1 \dddot{\hat x}\,dh=-\dddot{\hat x}h\Big|_0^1+\int_0^1 \hat x^{(4)}h\,dt=0.$$

Por consiguiente,

$$\mathcal{J}(x(\cdot)+h(\cdot))-\mathcal{J}(\hat x(\cdot))=\int_0^1 \ddot h^2\,dt\geqslant 0,$$

$$S_{\min}=\int_0^1 \ddot{\hat x}^2\,dt=\int_0^1 (6t-2)^2\,dt=4.$$

Resultado. La función $x=t^3-t^2$ proporciona al problema un mínimo absoluto, $S_{\text{mín}}=4$, $S_{\text{máx}}=+\infty$.

§ 6. Problemas de control óptimo

6.1. Principio de máximo de Pontriaguin

6.1.1. Planteamiento del problema. Llámase problema de *control óptimo* (en forma pontriaguiniana) el siguiente problema:

$$\mathcal{B}_0(\xi)\to\inf;\quad \mathcal{B}_i(\xi)\leqslant 0,\quad i=1,\ldots,m',$$
$$\mathcal{B}_i(\xi)=0,\quad i=m'+1,\ldots,m,$$
$$\Phi(\xi)\overset{\text{def}}{=}\dot x(t)-\varphi(t,x(t),u(t))=0\ \forall\, t\in T,\qquad (1)$$
$$u(t)\in U\ \forall\, t\in\Delta,\qquad (2)$$

donde $\xi=(x(\cdot),u(\cdot),t_0,t_1)$, $x(\cdot)\in KC^1(\Delta,\mathbf{R}^n)$, $u(\cdot)\in$ $\in KC(\Delta,\mathbf{R}^r)$, $t_0,t_1\in\operatorname{int}\Delta$, $t_0<t_1$, Δ es un segmento finito dado; U, un conjunto arbitrario de $\mathbf{R}^r$; y $T\subset\Delta$, un conjunto de puntos de continuidad de control $u(\cdot)$,

$$\mathcal{B}_i(\xi)=\int_{t_0}^{t_1} f_i(t,x(t),u(t))\,dt+\psi_i(t_0,x(t_0),$$
$$t_1,x(t_1)),\quad i=0,1,\ldots,m.$$

Señalemos que la ecuación (1), llamada *relación diferencial*, ha de cumplirse en todos los puntos de continuidad de la función vectorial $u(\cdot)=(u_1(\cdot),\ldots,u_r(\cdot))$. A diferencia del problema de Lagrange, aquí existe la acotación (2) de tipo de inclusión, mientras que la variable fásica $x(\cdot)=(x_1(\cdot),\ldots,x_n(\cdot))$ puede tener menor suavidad. El caso parcial (Pr) es un problema donde uno de los extremos del segmento de integración o incluso ambos extremos están fijos.

El elemento ξ para el cual se cumplen todas las condiciones y acotaciones mencionadas del problema, se llama *proceso admisible controlado*.

El proceso admisible controlado $\hat{\xi} = (\hat{x}(\cdot), \hat{u}(\cdot), \hat{t}_0, \hat{t}_1)$ se denomina (localmente) *óptimo (o proceso óptimo en el sentido fuerte)*, si existe $\delta > 0$ tal que para todo proceso admisible controlado $\xi = (x(\cdot), u(\cdot), t_0, t_1)$, con el cual

$$\| (x(\cdot), t_0, t_1) - (\hat{x}(\cdot), \hat{t}_0, \hat{t}_1 \| C(\Delta), \mathbf{R}^n)\, \mathbf{R}^2 < \delta,$$

se cumple la desigualdad $\mathscr{B}_0(\xi) \geqslant \mathscr{B}_0(\hat{\xi})$.

6.1.2. Formulación del problema en un caso general.

Teorema (principio de máximo de Pontriaguin). *Supongamos que* $\hat{\xi} = (\hat{x}(\cdot), \hat{u}(\cdot), \hat{t}_0, \hat{t}_1$ *es el proceso óptimo (en el sentido fuerte) de un problema de control óptimo* (Pr); *las funciones* f_i, $i = 0, 1, \ldots, m$, φ *y sus derivadas parciales en* x *son continuas en cierto entorno del conjunto* $\{(t, \hat{x}(t)) \mid t \in \Delta\}$ *multiplicado cartesianamente por* U; *y las funciones* ψ_i, $i = 0, 1, \ldots, m$ *son continuamente diferenciables en el entorno del punto* $(\hat{t}_0, \hat{x}(\hat{t}_0), \hat{t}_1, \hat{x}(\hat{t}_1))$ *(condición de suavidad).*

Entonces se hallarán los factores de Lagrange $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbf{R}^{m+1}$, $p(\cdot) \in \hat{KC}^1(\Delta, \mathbf{R}^n)$ *que no todos son iguales a cero, pero que son tales que para la función lagrangiana*

$$\mathscr{L} = \int_{t_0}^{t_1} [f(t, x, u) + p(t)(\dot{x} - \varphi(t, x, u))]\, dt + l(t_0, x(t_0), t_1, x(t_1)),$$

donde

$$f(t, x, u) = \sum_{i=0}^{m} \lambda_i f_i(t, x, u), \quad l = \sum_{i=0}^{m} \lambda_i \psi_i(t_0, x(t_0)\, t_1, x(t_1))$$

es un terminante, se cumplen las condiciones:

a) *de estacionaridad en* x, *es decir, la ecuación de Euler*

$$-\frac{d}{dt} \hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0 \Leftrightarrow \dot{p}(t) = \hat{f}_x(t) - p(t)\, \hat{\varphi}_x(t)$$

para la lagrangiana $L = f(t, x, u) + p(t)(\dot{x} - \varphi(t, x, u))$;

b) *de transversalidad en* x:

$$\hat{L}_{\dot{x}}(\hat{t}_0) = \hat{l}_{x(t_0)}, \quad p(\hat{t}_0) = \hat{l}_{x(t_0)},$$

$$\hat{L}_{\dot{x}}(\hat{t}_1) = -\hat{l}_{x(t_1)} p(\hat{t}_1) = -\hat{l}_{x(t_1)};$$

c) *de optimidad en* u:

$$\min_{\in U} L(t, \hat{x}(t), \dot{\hat{x}}(t), u) = L(t, \hat{x}(t), \dot{\hat{x}}(t), \hat{u}(t)) \Leftrightarrow$$

$$\Leftrightarrow \min_{u \in U} \{f(t, \hat{x}(t), u) - p(t)\, \varphi(t, \hat{x}(t), u)\} =$$

$$= f(t_1, \hat{x}(t), \hat{u}(t)) - p(t)\, \varphi(t, \hat{x}(t), \hat{u}(t)) \;\forall\; t \in T;$$

d) *de estacionaridad en t_0, t_1 (sólo se escribe para los extremos móviles del segmento de integración)*:

$$\hat{\mathcal{L}}_{t_0}=0, \qquad -\hat{f}(\hat{t}_0)+\hat{l}_{t_0}-\hat{l}_{x(t_0)}\dot{\hat{x}}(\hat{t}_0)=0;$$
$$\Longleftrightarrow$$
$$\hat{\mathcal{L}}_{t_1}=0 \qquad \hat{f}(\hat{t}_1)+\hat{l}_{t_1}+\hat{l}_{x(t_1)}\dot{\hat{x}}(\hat{t}_1)=0;$$

e) *de no rigidez complementaria*:

$$\lambda_i\mathcal{B}_i(\hat{\xi})=0, \quad i=1,\ \ldots,\ m';$$

f) *de no negatividad*:

$$\lambda_i\geqslant 0, \quad i=0,\ 1,\ \ldots,\ m'.$$

6.1.3. Formulación y demostración del principio de máximo de Pontriaguin en un caso parcial. Formulemos y demostremos el principio de máximo de Pontriaguin para el siguiente caso parcial del problema de control óptimo, o sea, para el problema de extremo libre y de tiempo fijo:

$$\mathcal{B}(x(\cdot),\ u(\cdot))=\int_{t_0}^{t_1} f(t,\ x(t),\ u(t))\,dt+\psi(x(t_1))\to\inf; \tag{Pr}$$

$$\dot{x}(t)=\varphi(t,\ x(t),\ u(t)),\ u(t)\in U,\ x(t_0)=x_0.$$

Teorema. *Supongamos que* $(\hat{x}(\cdot),\ \hat{u}(\cdot))$ *es un proceso óptimo (en el sentido fuerte) en el problema de control óptimo* (Pr); *las funciones* f, φ *y sus derivadas parciales son continuas en* x *en cierto entorno del conjunto* $\{(t,\ \hat{x}(t))\mid t\in[t_0,\ t_1]\}$ *multiplicado cartesianamente por* U; ψ *es continuamente diferenciable en el entorno del punto* $\hat{x}(t_1)$ (condición de *suavidad*); *y* U *es un conjunto arbitrario de* $\mathbf{R}^r$.

Entonces se cumple la condición de optimidad en u:

$$f(t,\ \hat{x}(t),\ u)-p(t)\,\varphi(t,\ \hat{x}(t),\ u)\geqslant\hat{f}(t)-p(t)\,\hat{\varphi}(t) \tag{1}$$
$$\forall t\in T, \quad \forall u\in U,$$

donde $p(\cdot)$ *es la única solución de la ecuación diferencial*

$$\dot{p}(t)+p(t)\,\hat{\varphi}_x(t)=\hat{f}_x(t)\ \forall\ t\in T \tag{2}$$

con condición de contorno

$$p(t_1)=-\psi'(\hat{x}(t_1)). \tag{3}$$

Cabe señalar que el principio de optimidad (1) con condiciones (2) – (3) puede ser fácilmente deducido de las condiciones necesarias de optimidad en el problema general de control óptimo. El factor de Lagrange λ_0 con la funcional $\mathcal{B}$ resulta igual a la unidad y, en este caso, la condición de transversabilidad en $x(t_0)$ no tiene importancia.

◁ La unicidad de solución de la ecuación (2) con la condición de contorno (3) se desprende del teorema de existencia y unicidad de la solución del problema de Cauchy para los sistemas lineales [ATF, pág. 191].

A) **Variaciones aciculares.** Fijemos el punto $\tau \in T$, el elemento $v \in U$ y el número $\alpha \geqslant 0$ tan pequeño que $[\tau - \alpha, \tau] \subset T$.

Llamemos el control

$$u_\alpha(t) = \begin{cases} \hat{u}(t) & t \notin [\tau-\alpha, \tau) \\ v & t \in [\tau-\alpha, \tau) \end{cases}$$

variación elemental (acicular) de control $\hat{u}$.

Sea $x_\alpha(\cdot)$ la solución de la ecuación $\dot{x}(t) = \varphi(t, x(t), u_\alpha(t)$ con condición inicial $x(t_0) = x_0$. Según el teorema local de existencia [ATF, págs. 186—189], la función $x_\alpha(\cdot)$ se determina cuando $\alpha \leqslant \alpha_0$ en cierto entorno del punto t_0, pero del lema 1 formulado más abajo se desprende que, en realidad, la función vectorial $x_\alpha(\cdot)$ puede ser determinada de una sola manera en todo el segmento $[t_0, t_1]$. La función vectorial $x_\alpha(\cdot)$ se llama *variación elemental (acicular) de la función* $\hat{x}(\cdot)$, mientras que el par $(\hat{x}_\alpha(\cdot), \hat{u}_\alpha(\cdot)$ se denomina *variación elemental del proceso* $(\hat{x}(\cdot), \hat{u}(\cdot))$. El par (τ, v) que determina esa variación se llama *aguja elemental*.

B) **Lema 1** (acerca de las propiedades de variación elemental). *Sea* (τ, v) *una aguja elemental fija. Entonces existirá un número* $\alpha_0 > 0$ *tal que* $[\tau - \alpha_0, \tau] \subset T$, *y la función* $x_\alpha(\cdot)$ *será definida, para cualquier* $\alpha \in [0, \alpha_0]$, *en todo el segmento* $[t_0, t_1]$. *Además,* $x_\alpha(\cdot) \to \hat{x}(\cdot)$ en la métrica del espacio $C([t_0, t_1], \mathbf{R}^n)$,

$$x_\alpha(t) - \hat{x}(t) = \alpha y(t) + r_\alpha(t) \quad \forall\, t \in [\tau, t_1],$$

en este caso $\sup\limits_{t \in [\tau, t]} \dfrac{|r_\alpha(t)|}{\alpha} \to 0$ *cuando* $\alpha \to +0$, *donde la función* $y(\cdot)$ *es continuamente diferenciable a trozos en* $[\tau, t_1]$ *y satisface la ecuación diferencial*

$$\dot{y}(t) = \hat{\varphi}_x'(t)\, y(t) \quad \forall\, t \in [\tau, t_1] \cap T \tag{4}$$

cuya condición inicial es

$$y(\tau) = \varphi(\tau, \hat{x}(\tau), v) - \hat{\varphi}(\tau) = \Delta\varphi(\tau, v).$$

La demostración del lema se basa en dos hechos fundamentales de la teoría de ecuaciones diferenciales: en el teorema local de existencia y en el teorema de dependencia continua y diferenciabilidad de la solución a base de los datos iniciales. Aquí no exponemos esa demostración, ya que la misma se ofrece en el libro [ATF, págs. 89—91].

C) **Lema 2** (acerca del incremento de la funcional). *Sea* (τ, υ) *una aguja elemental fija*, $\chi(\alpha) = \mathscr{B}(x_\alpha(\cdot), u_\alpha(\cdot))$. *Entonces la función* $\chi(\cdot)$ *será diferenciable a la derecha en cero y*

$$\chi'(+0) = f(\tau), \quad \hat{x}(\tau, \upsilon) - \hat{f}(\tau) - p(\tau)\, y(\tau).$$

◁◁ Utilizando el teorema del valor medio para las funciones numéricas, la regla de diferenciación bajo signo integral, y el lema 1, obtenemos

$$\chi'(+0) = \lim_{\alpha \to +0} \frac{\chi(\alpha) - \chi(0)}{\alpha} = \lim_{\alpha \to +0} \frac{1}{\alpha}\left(\int_{t_0}^{t_1} f(t, x_\alpha(t), u_\alpha(t))\, dt - \right.$$

$$\left. - \int_{t_0}^{t_1} f(t, \hat{x}(t), \hat{u}(t))\, dt\right) + \lim_{\alpha \to +0} \frac{\psi(x_\alpha(t_1)) - \psi(\hat{x}(t_1))}{\alpha} =$$

$$= \lim_{\alpha \to +0} \frac{1}{\alpha}\left(\int_{\tau-\alpha}^{\tau} \left(f(t, x_\alpha(t), \vartheta) - \hat{f}(t)\right) dt + \right.$$

$$\left. + \int_{\tau}^{t_1} (f(t, x_\alpha(t), \hat{u}(t)) - \hat{f}(t))\, dt\right) + \psi'(\hat{x}(t_1))\, y(t_1) =$$

$$= f(\tau, \hat{x}(\tau), \vartheta) - \hat{f}(\tau) + \int_{\tau}^{t_1} \hat{f}_x(t)\, y(t)\, dt + \psi'(\hat{x}(t_1))\, y(t_1).$$

Expresando $\hat{f}_x$ de la ecuación (2) y teniendo en cuenta la ecuación (4), obtenemos

$$\int_{\tau}^{t_1} \hat{f}_x(t)\, y(t)\, dt = \int_{\tau}^{t_1} (\dot{p}(t) + p(t)\,\hat{\varphi}_x(t))\, y(t)\, dt =$$

$$= \int_{\tau}^{t_1} (\dot{p}(t)\, y(t) + p(t)\, \dot{y}(t))\, dt = \int_{\tau}^{t_1} \frac{d}{dt}(p(t)\, y(t))\, dt =$$

$$= p(t_1)\, y(t_1) - p(\tau)\, y(\tau).$$

Introduciendo el valor obtenido $\int_{\tau}^{t_1} \hat{f}_x y\, dt$ en la expresión de $\chi'(+0)$ y teniendo en cuenta la condición 3), obtenemos la representación buscada. ▷▷

D) **Terminación de la demostración.** Del lema 1 se deduce que si $\alpha \in [0, \alpha_0]$, en este caso $(x_\alpha(\cdot), u_\alpha(\cdot))$ será un proceso admisible controlado y $x_\alpha(\cdot)$ tenderá uniformemente a $\hat{x}(\cdot)$. Como $(\hat{x}(\cdot), \hat{u}(\cdot))$ es un proceso óptimo, entonces, para pequeños valores de

$\alpha > 0$,

$$\mathscr{B}(x_\alpha(\cdot), u_\alpha(\cdot)) \geqslant \mathscr{B}(\hat{x}(\cdot), \hat{u}(\cdot)) \Leftrightarrow \chi(\alpha) \geqslant \chi(0).$$

De aquí, según el lema $2\chi'(+0) \geqslant 0$ y de la expresión para $\chi'(+0)$ e $y(\tau)$, se deduce que

$$f(\tau, \hat{x}(\tau), \upsilon) - p(\tau)\varphi(\tau, \hat{x}(\tau), \upsilon) \geqslant f(\tau) - p(\tau)\hat{\varphi}(\tau)$$
$$\forall \tau \in T, \quad \forall \upsilon \in U,$$

es decir, se cumple la relación (1). ▷

6.1.4. Demostración del principio de máximo de Pontriaguin en un caso general.

Expongamos en los puntos A) — D) una serie de afirmaciones y construcciones auxiliares.

A) **Lema sobre el sistema centrado.** *Supongamos que K es un compacto, y $\{K_\alpha\}_{\alpha \in \mathfrak{U}}$, un sistema de subconjuntos cerrados de K, cuyo cualquier subsistema finito tiene una intersección no vacía (sistema centrado) Entonces, la interseccion de todos los conjuntos del sistema $\{K_\alpha\}_{\alpha \in \mathfrak{U}}$ tampoco será vacía.*

◁◁ Denotemos por $\mathcal{O}_\alpha$ el complemento a K_α en K ($\mathcal{O}_\alpha = K \setminus K_\alpha$). En este caso $\mathcal{O}_\alpha$ estará abierto en K. Si $\bigcap\limits_{\alpha \in \mathfrak{U}} K_\alpha = \varnothing$, entonces $\bigcup\limits_{\alpha \in \mathfrak{U}}^{m} \mathcal{O}_\alpha = K$, es decir, $\{\mathcal{O}_\alpha\}_{\alpha \in \mathfrak{U}}$ será el recubrimiento abierto del compacto K. Partiendo de la definición del compacto podemos hallar $\alpha_1, \ldots, \alpha_m$ tales que $\bigcup\limits_{i=1}^{m} \mathcal{O}_{\alpha_i} = K$. Pero entonces $\bigcup\limits_{i=1}^{m} K_{\alpha_i} = \varnothing$, en contradicción con la definición del sistema centrado. Eso significa que $\bigcap\limits_{\alpha \in \mathfrak{U}} K_\alpha \neq \varnothing$. ▷▷

B) **Variaciones aciculares y paquete de agujas.** Sometamos a variación el proceso $\hat{\xi}$ incluyéndolo en una familia finitamente paramétrica, determinada por un paquete de agujas (conjunto de variaciones aciculares de control $\hat{u}(\cdot)$). Para eso fijemos el número natural N, o sea, los conjuntos: de puntos $\tau = (\tau_1, \ldots, \tau_N)$, $\tau_1 \leqslant \tau_2 \leqslant \ldots \leqslant \tau_N$, de controles $\upsilon = (\upsilon_1, \ldots, \upsilon_N)$, y de longitudes $\alpha = (\alpha_1, \ldots, \alpha_N)$ $(\tau_i \in T,\ \upsilon_i \in U,\ \alpha_i \geqslant 0,\ i = 1, \ldots, N)$. Denotemos por $|\alpha|$, como siempre, $\left(\sum\limits_{i=1}^{n} \alpha_i^2\right)^{1/2}$.

$$u_\alpha(t) = \begin{cases} \hat{u}(t), & t \in \Delta \setminus \bigcup\limits_{i=1}^{N} \Delta_i; \\ \vartheta_i, & t \in \Delta_i, \end{cases}$$

donde $\Delta_i = [\tau - (N - i)|\alpha| - \alpha_i,\ \tau_i - (N - i)|\alpha|)$ se denomina *variación acicular de control* $\hat{u}(\cdot)$ y se halla determinada por el paquete de agujas (τ, υ, α). Algunos puntos τ_i pueden coincidir. No obstante, los semiintervalos Δ_i (de longitud α_i) se eligen de tal modo que no se intersequen y que, con pequeño $|\alpha|$, permanezcan en el conjunto T.

Sea $x_\eta(\cdot)$ $(\eta = (t_0, x_0, \alpha))$ la solución de la ecuación

$$\dot{x} = \varphi(t, x, u_\eta(t)) \tag{1}$$

cuya condición inicial es $x(t_0) = x_0$ y donde el punto (t_0, x_0) se halla en el entorno δ del punto $(\hat{t}_0, \hat{x}_0)$ $(\hat{x}_0 = \hat{x}(\hat{t}_0))$. Más abajo demostraremos que $x_\eta(\cdot)$, es decir, que la solución de la ecuación diferencial existe realmente y que la misma se halla determinada en todo el segmento Δ.

La función vectorial $x_\eta(\cdot)$ se llama *variación acicular de la función* $\hat{x}(\cdot)$ determinada por el punto (t_0, x_0) y el paquete de agujas (τ, υ, α).

C) **Acerca de los sistemas de ecuaciones diferenciales.** *Supongamos que el problema de Cauchy en el segmento finito* Δ

$$\dot{x} = F(t, x), \ x(t_0) = x_0 \ (t_0 \in \operatorname{int} \Delta)$$

tiene la solución $\hat{x}(\cdot) \in KC^1(\Delta, \mathbf{R}^n)$; *además,* F *es continua y continuamente diferenciable en* x *en cierto entorno* G *de la trayectoria* $\Gamma_{\hat{x}} = \{(t, \hat{x}(t)) \mid t \in \Delta\}$.

Entonces se hallará $G' \subset G$, *es decir, un entorno de la trayectoria* $\Gamma_{\hat{x}}$, *tal que para cualquier punto* $(t_0, x_0) \in G'$ *existirá la única solución* $x(\cdot, t_0, x_0)$ *del problema de Cauchy, determinada en* Δ; *además, la función* $x(t, t_0, x_0)$ *es continuamente diferenciable en el entorno* $\Delta \times G'$ *y*

$$\left.\frac{\partial x(t, t_0, x_0)}{\partial x_0}\right|_{x_0 = \hat{x}(t_0)} = \Omega(t, t_0),$$

$$\left.\frac{\partial x(t, t_0, x_0)}{\partial t_0}\right|_{x_0 = \hat{x}(t_0)} = -\Omega(t, t_0) F(t_0, \hat{x}(t_0)),$$

donde $\Omega(t, t_0)$ *es el sistema fundamental de soluciones de la ecuación en variaciones*:

$$\frac{d}{dt}\Omega(t, t_0) = F_x(t, \hat{x}(t))\,\Omega(t, t_0),$$

$$\Omega(t_0, t_0) = 1,$$

donde I *es la matriz unidad.*

Se trata del teorema clásico de existencia y dependencia continuamente diferenciable de la solución de la ecuación diferencial a partir de los datos iniciales [ATF, págs. 195—204].

D) **Lema de la variación acicular.** *Sean fijos los conjuntos* τ *y* υ *en el paquete de agujas* (τ, υ, α). *En este caso existirá* $\varepsilon > 0$ *tal que si*

$0 < |\alpha| < \varepsilon$, $|x_0 - \hat{x}_0| < \varepsilon$, $|t_0 - \hat{t}_0| < \varepsilon$, *entonces* $\Delta_i \subset T$ *y, además,*

1. *La función* $x_\eta(\cdot)$, *es decir, la solución de la ecuación* (1) *se hallará determinada en el segmento* Δ, $\|x_\eta(\cdot) - \hat{x}(\cdot)\|_{C(\Delta, R^n)} \to 0$ *cuando* $\eta \to \hat{\eta} = (\hat{t}_0, \hat{x}_0, 0)$.

2. *La aplicación* $(t, \eta) \to x_\eta(t)$ *continuará hasta alcanzar la aplicación continuamente diferenciable en cierto entorno del punto* $(\hat{t}_1, \hat{\eta})$.

3. *Se cumplirán las condiciones*

$$\left.\frac{\partial x_\eta(t)}{\partial x_0}\right|_{\eta=\hat{\eta}} = \Omega(t, \hat{t}_0),$$

$$\left.\frac{\partial x_\eta(t)}{\partial t_0}\right|_{\eta=\hat{\eta}} = \Omega(t, \hat{t}_0)\,\hat{\varphi}(\hat{t}_0),$$

donde $\Omega(t, t_0)$ *es el sistema fundamental de soluciones de la ecuación en variaciones*:

$$\Omega(t, t_0) = \hat{\varphi}_x(t)\,\Omega(t, t_0), \quad \Omega(t_0, t_0) = I.$$

Tracemos el método de demostración del lema. Si el control $\hat{u}(\cdot)$ es una función continua, las afirmaciones del lema serán inmediatamente deducidas del punto C) del teorema sobre la existencia y dependencia continuamente diferenciable de la solución de la ecuación diferencial partiendo de los datos iniciales. Pero si $\hat{u}(\cdot)$ es continua a trozos, entonces será necesario aplicar el teorema C) varias veces en cada sector de continuidad.

E) **Reducción a un problema de dimensión finita.** Fijemos otra vez N, τ y υ. Efectuemos la referida reducción. Denotemos $z = (t_1, \eta) = (t_1, t_0, x_0, \alpha_1, \ldots, \alpha_N) \in \mathbf{R}^{2+n+N}$, supongamos que

$$F_i(z) = \mathscr{B}_i(x_\eta(\cdot), u_\alpha(\cdot), t_0, t_1) =$$
$$= \int_{t_0}^{t_1} f_i(t, x_\eta(t), u_\alpha(t))\,dt + \psi_i(t_0, x_0, t_1, x_\eta(t_1))$$

y examinemos el problema

$$F_0(z) \to \inf; \quad F_i(z) \leqslant 0, \quad i = 1, \ldots, m',$$
$$F_i(z) = 0,$$
$$i = m' + 1, \ldots, m, \quad \alpha_j \geqslant 0, \quad j = 1, \ldots, N. \qquad (\mathrm{Pr}_{\tau,\nu})$$

En virtud del lema acerca de la variación acicular, F_i son continuamente diferenciables en cierto entorno del punto $\hat{z} = (\hat{t}_1, \hat{t}_0, \hat{x}_0, 0)$ y $(x_\eta(\cdot), t_0, t_1) \to (\hat{x}(\cdot), \hat{t}_0, \hat{t}_1)$ cuando $z \to \hat{z}$. Pero como $\hat{\xi}$ proporciona un mínimo local al problema (Pr), entonces el punto

$\hat{z} \in \operatorname{locmín} \mathrm{Pr}_{\tau,\nu}$. Eso significa que al problema $\mathrm{Pr}_{\tau,\eta}$ se le puede aplicar el principio de Lagrange para los problemas de dimensión finita con igualdades y desigualdades (p. 2.4). Conforme a dicho principio se hallarán los factores lagrangianos $\bar{\lambda}_0, \ldots, \bar{\lambda}_m, \bar{\mu}_1, \ldots, \bar{\mu}_\nu$ que no todos son iguales a cero ($\bar{\lambda}_i = \bar{\lambda}_i(\tau, \upsilon)$, $\bar{\mu}_j = \bar{\mu}_j(\tau, \upsilon)$), pero que son tales que para la función lagrangiana

$$\mathscr{L} = \sum_{i=0}^{m} \bar{\lambda}_i F_i(z) - \sum_{j=1}^{N} \bar{\mu}_j \alpha_j$$

se cumplen las condiciones de estacionaridad ($\bar{\mathscr{L}}_z = 0$), de no negatividad ($\bar{\lambda}_i \geqslant 0$, $i = 0, 1, \ldots, m'$, $\bar{\mu}_j \geqslant 0$, $j = 1, \ldots, N$) y de no rigidez complementaria ($\bar{\lambda}_i F_i(\hat{z}) = 0$, $i = 1, \ldots, m'$, $\mu_j \alpha_j = 0$, $j = 1, \ldots, N$).

F) **Transformación de las condiciones necesarias del problema de dimensión finita.** Supongamos que

$$\bar{f}(t, x, u) = \sum_{i=0}^{m} \bar{\lambda}_i f_i(t, x, u),$$

$$\bar{l}(t_0, x_0, t_1, x_1) = \sum_{i=0}^{m} \bar{\lambda}_i \psi_i(t_0, x_0, t_1, x_1),$$

donde $\bar{p}(\cdot)$ es la solución de la ecuación diferencial

$$\dot{\bar{p}}(t) + \bar{p}(t) \hat{\varphi}_x(t) = \hat{\bar{f}}_x(t) \tag{2}$$

con la condición de contorno

$$\bar{p}(\hat{t}_1) = -\hat{\bar{l}}_{x_1}. \tag{3}$$

De esas determinaciones y de la determinación de $\Omega(t, \hat{t}_0)$ se deduce que

$$\frac{d}{dt}(p(t)\Omega(t, \hat{t}_0)) = \hat{\bar{f}}_x(t)\Omega(t, \hat{t}_0) \quad \text{y}$$

$$\bar{\mathscr{L}} = \int_{t_0}^{t_1} \bar{f}(t, x_\eta, u_{\bar{\alpha}})\, dt + \bar{l}(t_0, x_0, x_\eta(t_1)) - \sum_{1}^{N} \bar{\mu}_j \alpha_j.$$

Escribamos las condiciones de estacionaridad de la función de Lagrange $\bar{\mathscr{L}}$ en el punto $\hat{z}$ teniendo en cuenta el lema de incremento de la funcional (p. 6.1.3) y las fórmulas del lema de la variación acicular:

$$\frac{\partial \overline{\mathcal{L}}(\hat{z})}{\partial \alpha_k} = \Delta \bar{f}(\tau_k,\ \vartheta_k) - \bar{p}(\tau_k)\,\Delta\varphi(\tau_k,\ \vartheta_k) - \bar{\mu}_k = 0,\ k = 1,\ \ldots,\ N, \quad (4)$$

$$\frac{\partial \overline{\mathcal{L}}(\hat{z})}{\partial x_0} = \int_{t_0}^{\hat{t}_1} \hat{\bar{f}}_x(t)\,\frac{\partial x_\eta(t)}{\partial x_0}\Bigg|_{\eta=\hat{\eta}} dt + \hat{\bar{l}}_{x_0} +$$

$$+ \hat{\bar{l}}_{x_1}\,\frac{\partial x_\eta(\hat{t}_1)}{\partial x_0}\Bigg|_{\eta=\hat{\eta}} = \bar{p}(\hat{t}_1)\,\Omega(\hat{t}_1,\ \hat{t}_0) - \bar{p}(\hat{t}_0) + \hat{\bar{l}}_{x_0} + \hat{\bar{l}}_{x_1}\Omega(\hat{t}_1,\ \hat{t}_0) =$$

$$= -\bar{p}(\hat{t}_0) + \hat{\bar{l}}_{x_0} = 0, \quad (5)$$

$$\frac{\partial \overline{\mathcal{L}}(\hat{z})}{\partial t_0} = -\hat{\bar{f}}(\hat{t}_0) + \int_{\hat{t}_0}^{\hat{t}_1} \hat{\bar{f}}_x(t)\,\frac{\partial x_\eta(t)}{\partial t_0}\Bigg|_{\eta=\hat{\eta}} dt + \hat{\bar{l}}_{t_0} +$$

$$+ \hat{\bar{l}}_{x_1}\,\frac{\partial x_\eta(\hat{t}_1)}{\partial t_0}\Bigg|_{\eta=\hat{\eta}} - \hat{\bar{f}}(\hat{t}_0) - \bar{p}(\hat{t}_1)\,\Omega(\hat{t}_1,\ \hat{t}_0)\,\dot{\hat{x}}(t_0) + \bar{p}(\hat{t}_0)\,\dot{\hat{x}}(\hat{t}_0) +$$

$$+ \hat{\bar{l}}_{t_0} + \hat{\bar{l}}_{x_1}\Omega(\hat{t}_1,\ \hat{t}_0)\,\dot{\hat{x}}(\hat{t}_0) = -\hat{\bar{f}}(\hat{t}_0) +$$

$$+ \bar{p}(\hat{t}_0)\,\dot{\hat{x}}(\hat{t}) + \hat{\bar{l}}_{t_0} \overset{(5)}{=} -\hat{\bar{f}}(\hat{t}_0) + \hat{\bar{l}}_{t_0} + \hat{\bar{l}}_{x_0}\dot{\hat{x}}(\hat{t}) = 0, \quad (6)$$

$$\frac{\partial \overline{\mathcal{L}}(\hat{z})}{\partial t_1} = \hat{\bar{f}}(\hat{t}_1) + \hat{\bar{l}}_{t_1} - \hat{\bar{l}}_{x_1}\dot{\hat{x}}(\hat{t}_1) = 0. \quad (7)$$

Evidentemente que $\bar{\lambda} \neq 0$, de lo contrario, de las definiciones de $\bar{f}$ y $\bar{p}$, así como de las relaciones (4) resulta que $\bar{\mu} = 0$, lo cual es imposible. Multiplicando por una constante positiva normalizamos el vector $\bar{\lambda}$ de tal modo que $\sum_0^m \lambda_i^2 = 1$.

Así queda demostrado que existe un vector unidad $\bar{\lambda}$ tal que se cumplen las relaciones (2) — (7) y, además,

$$\bar{\lambda}_i F_i(z) = 0 \quad \bar{\lambda}_i \mathscr{B}_i(\hat{\xi}) = 0,\quad i = 1,\ \ldots,\ m', \quad (8)$$

$$\bar{\lambda}_i \geqslant 0,\quad i = 0,\ 1,\ \ldots,\ m'. \quad (9)$$

G) **Terminación de la demostración.** Examinemos, en el espacio $\mathbf{R}^{m+1}$, los subconjuntos $K(\tau,\ \upsilon)$, $\upsilon \in U$, $\tau \in T$ de la esfera compacta $S = \{\lambda \mid \sum_0^m \lambda_i^2 = 1\}$, que constan de los vectores λ para los cuales se cumplen las afirmaciones a) — f) del teorema acerca del principio de máximo de Pontriaguin tomando $t = \tau$, $u = \upsilon$ en el p. C). De la definición del conjunto $K(\tau,\ \upsilon)$ se deduce fácilmente que los subconjuntos están cerrados. En virtud de las condiciones 2—9, cualquier intersección finita de los conjuntos $K(\tau_k, \upsilon_k)$, $k = 1, \ldots, N$ no es una intersección vacía. Según el lema del sistema centrado, todos los conjuntos $K(\tau, \upsilon)$ tienen una intersección no vacía. Eso significa que existe un vector no nulo $\lambda = (\lambda_0, \ldots, \lambda_m)$ y una función $p(\cdot) \in KC^1(\Delta, \mathbf{R}^n)$ tales que se cumplen las afirmaciones del teorema a) — f) con una condición de optimidad que se cumple para cualquier $\tau \in T$, $\upsilon \in U$.

Nota. El principio de máximo está demostrado en el espacio

$$KC^1(\Delta, \mathbf{R}^n) \times KC(\Delta, \mathbf{R}^r) \times \mathbf{R}^2.$$

Cambios insignificantes de la demostración permiten argumentarla en el espacio

$$W^1_\infty(\Delta, \mathbf{R}^n) \times L_\infty(\Delta, \mathbf{R}^r) \times \mathbf{R}^2.$$

Además de la obra clásica de L. S. Pontriaguin, también puede ser recomendada la monografía de R. V. Gamkrelidze «Fundamentos del control óptimo» (Tbilisi, Edit. de la Universidad de Tbilisi, 1977, en ruso).

6.2. Ejemplos

Ejemplo 1.

$$\mathscr{B}(x(\cdot)) = \int_0^4 (\dot{x}^2 + x)\,dt \to \inf; \quad |\dot{x}| \leqslant 1, \quad x(0) = 0.$$

Solución. Reduzcamos el problema al tipo de problemas de control óptimo introduciendo el control $u(\cdot)$:

$$\int_0^4 (u^2 + x)\,dt \to \inf; \quad \dot{x} = u, \quad u \in [-1, 1], \quad x(0) = 0.$$

Función de Lagrange:

$$\mathscr{L} = \int_0^4 (\lambda_0(u^2 + x) + p(\dot{x} - u))\,dt - \lambda x(0).$$

Condiciones necesarias:

a) ecuación de Euler para la lagrangiana $L = \lambda_0(u^2 + x) + p(\dot{x} - u)$:

$$-\frac{d}{dt} L_{\dot{x}} + L_x = 0 \Longleftrightarrow \dot{p} = \lambda_0;$$

b) de transversalidad en x:

$$L_{\dot{x}}(t_k) = (-1)^k l_{x(t_k)}, \quad k = 0 \Longleftrightarrow p(0) = \lambda,$$
$$p(4) = 0;$$

c) de optimidad en u:

$$\min_{u \in [-1,\ 1]} (\lambda_0 u^2 - pu) = \lambda_0 \hat{u}^2 - p\hat{u}.$$

Si $\lambda_0 = 0$, en este caso, de a) $\dot{p} = 0$ y de b) $p = \lambda = 0$ resulta que todos los factores lagrangianos son ceros. Supongamos que $\lambda_0 = 1$. Entonces, de

a) $\dot{p} = 1$ y de b) $p = t - 4$. De la condición c) se deduce que

$$\hat{u} = \begin{cases} \operatorname{sign} p, & |p/2| > 1; \\ p/2, & |p/2| \leqslant 1, \end{cases} \Leftrightarrow \dot{\hat{x}} = \begin{cases} -1, & 0 \leqslant t \leqslant 2 \\ (t-4)/2 & 2 \leqslant t \leqslant 4; \end{cases}$$

De la condición inicial hallamos la función continua

$$\hat{x}(t) = \begin{cases} -t & 0 \leqslant t \leqslant 2; \\ t^2/4 - 2t + 1, & 2 \leqslant t \leqslant 4. \end{cases}$$

Demostremos, a base de verificación directa, que $\hat{x}(\cdot) \in \operatorname{absmín}$. Tomamos una función $h(\cdot) \in KC^1([0, 4])$ tal que $\hat{x}(\cdot) + h(\cdot)$ sea admisible en el problema. Para esto es necesario adoptar una función $h(\cdot)$ tal para la cual $|\dot{\hat{x}} + \dot{h}| \leqslant 1$, $h(0) = 0$. Integrando por partes y teniendo en cuenta que $\dot{\hat{x}}(t) = (t-4)/2$ para $2 \leqslant t \leqslant 4$, obtenemos

$$\mathscr{B}(\hat{x}(\cdot) + h(\cdot)) - \mathscr{B}(\hat{x}(\cdot)) = \int_0^4 ((\dot{\hat{x}} + \dot{h})^2 + \hat{x} + h)\, dt -$$

$$- \int_0^4 (\dot{\hat{x}}^2 + \hat{x})\, dt = \int_0^4 2\dot{\hat{x}}\dot{h}\, dt + \int_0^4 h\, dt + \int_0^4 \dot{h}^2\, dt \geqslant 2\int_0^4 \dot{\hat{x}}\, dt + \int_0^4 h\, dt =$$

$$= 2\dot{\hat{x}}(t)\, h(t)\Big|_0^4 + \int_0^4 (-2\ddot{\hat{x}} + 1)\, h\, dt = \int_0^2 h\, dt \geqslant 0,$$

ya que $h(t) \geqslant 0$ para $t \in [0, 2]$ cuando $h(0) = 0$ y $\dot{h}(t) \geqslant 0$ cuando $t \in [0, 2]$. Así, $\hat{x} \in \operatorname{absmín}$.

Ejemplo 2 (problema elemental de acción rápida). Supongamos que tenemos un carro que se mueve rectamente, sin rozamiento, por rieles horizontales. El mismo es manejado (controlado) por una fuerza exterior que puede variar dentro de límites programados. Es necesario parar el carro en un lugar determinado durante un tiempo mínimo.

Denotemos por $x(t)$ la coordenada corriente del carro; por $x(0) = \xi_1$, su coordenada inicial; y por $\dot{x}(0) = \xi_2$, su velocidad inicial. Para simplificar el problema, supongamos que la masa del carro es $m = 1$. Entonces, según la ley de Newton, $\ddot{x} = u$. Establezcamos la limitación de la tracción en forma de $u \in [-1, 1]$. Es preciso parar el carro ($\dot{x}(T) = 0$) al comienzo de las coordenadas ($x(T) = 0$). Obtenemos la siguiente formalización del problema:

$$T \to \inf; \quad |\ddot{x}| \leqslant 1, \quad x(0) = \xi_1, \quad \dot{x}(0) = \xi_2, \quad x(T) = \dot{x}(T) = 0.$$

Solución. Reduzcamos el problema al tipo de problemas de control óptimo (p. 6.1.1), sustituyendo las variables $x_1 = x$, $x_2 = \dot{x}$ e introduciendo el control $u = \ddot{x}$:

$$T \to \inf; \quad \dot{x}_1 = x_2, \quad \dot{x}_2 = u, \quad u \in [-1, 1],$$
$$x_1(0) = \xi_1, \quad x_2(0) = \xi_2, \quad x_1(T) = x_2(T) = 0.$$

Función de Lagrange:

$$\mathscr{L} = \int_0^T (p_1)(\dot{x} - x_2) + p_2(\dot{x}_2 - u))\,dt + \lambda_0 T +$$

$$+ \lambda_1 (x_1(0) - \xi_1) + \lambda_2 (x_2(0) - \xi_2) +$$

$$+ \lambda_3 x_1(T) + \lambda_4 x_2(T).$$

Condiciones necesarias.

a) Ecuación de Euler para la lagrangiana $L = p_1(\dot{x} - x_2) + p_2(\dot{x}_2 - u)$:

$$-\frac{d}{dt} L_{\dot{x}_i} + L_{x_i} = 0, \quad i = 1,\ 2, \Longleftrightarrow \dot{p}_1 = 0,$$

$$\dot{p}_2 = -p_1 \Rightarrow p_2(t) = Ct + C_1.$$

b) Transversalidad en x para el terminante $l = \lambda_0 T + \lambda_1 (x_1(0) - \xi_1) + \lambda_2 (x_2(0) - \xi_2) + \lambda_3 x_1(T) - \lambda_4 x_2(T)$:

$$p_1(0) = \lambda_1 \quad p_2(0) = \lambda_2, \quad p_1(T) = -\lambda_3,$$

$$P_2(T) = -\lambda_4.$$

c) Optimidad en u (no se escriben los sumandos que no dependen de u):

$$\min_{u\in[-1,1]} \{-p_2(t)\,u\} = -p_2(t)\,\hat{u}(t) \Rightarrow \hat{u}(t) = \operatorname{sign} p_2(t) \text{ para } p_2(t) \neq 0.$$

d) estacionaridad en T.

$$\mathscr{L}_T = 0 \Longleftrightarrow \lambda_0 - \lambda_3 \dot{x}_1(T) + \lambda_4 \dot{x}_2(T) = 0.$$

Tomando en consideración que

$$\dot{x}_1(T) = 0, \quad \lambda_4 = -p_2(T), \quad p_2(T)\,\hat{u}(T) = |\,p_2(T)\,|,$$

obtenemos $\mathscr{L}_T = 0 \Longleftrightarrow \lambda_0 = |\,p_2(T)\,|$.

Si $\lambda_0 = 0$, de d) se desprende que $p_2(T) = 0$. En este caso p_2 no puede ser un cero idéntico, ya que de lo contrario todos los factores lagrangianos también serían ceros. Por lo tanto, de a) se desprende que $p_2(t) = C(t - T)$, pero en este caso, de c) se deduce que $\hat{u}(t) \equiv 1$ ó $\hat{u}(t) \equiv -1$. El conjunto de condiciones iniciales, correspondientes a tales controles, se describe por la ecuación

$$\xi_2 = \varphi(\xi_1), \quad \varphi(\xi_1) = \begin{cases} -\sqrt{2\xi_1}, & \xi_1 \geqslant 0; \\ \sqrt{-2\xi_1}, & \xi_1 \leqslant 0 \end{cases}$$

($\hat{u}(t) \equiv 1 \Rightarrow x_2(t) = t - T,\ x_1(t) = (t - T)^2/2 \Rightarrow \xi_{2^2/2} = \xi_1$; (el caso de $\hat{u} \equiv -1$ es análogo).

Pero si $\xi_2 \neq \varphi(\xi_1)$, entonces $\lambda_0 \neq 0$ y suponemos que $\lambda_0 = 1$. En este caso, de d) se deduce que $|\,p_2(T)\,| = 1$, o sea, existen dos posibilidades:

$$p_2^+(t) = C(t - T) + 1, \quad p_2^-(t) = C(t - T) - 1.$$

A estas posibilidades, en virtud de c) les corresponden los siguientes controles:

$$u^+(t) = \begin{cases} -1, & 0 \leqslant t < \tau; \\ 1, & \tau < t \leqslant T, \end{cases} \qquad u^-(t) = \begin{cases} 1, & 0 \leqslant t < \tau; \\ -1, & \tau < t \leqslant T \end{cases}$$

Examinemos las trayectorias correspondientes a los controles óptimos u^+ y u^- en el plano (x_1, x_2) llamado *plano fásico* (*fig.* 3).

Para los valores de t que les corresponde $\hat u(t) = 1$, tenemos

$$\dot x_2 = 1 \Rightarrow \dot x_1 = x_2 = t + C' \Rightarrow x_1 = t^2/2 + C't + C'' = x_2^2/2 + C.$$

Por consiguiente, la trayectoria fásica que corresponde a esos valores es un trozo de la parábola $x_1 = x_2^2/2 + C$. El sentido del movimiento a lo largo de esa parábola se halla determinado por el incremento de x_2, ya que en este

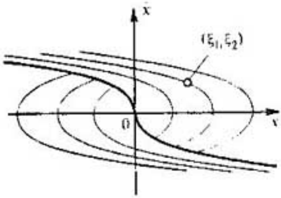


Fig 3

caso $\dot x_2 = 1$. Análogamente obtenemos que para los valores de t, la trayectoria fásica $\hat u(t) = -1$ es el trozo de parábola $x_1 = -x_2^2/2 + C$, mientras que el sentido del movimiento se deduce de la condición de disminución de x_2, ya que $\dot x_2 = -1$.

Ahora señalemos el lugar en el plano fásico (x_1, x_2) donde ha de realizarse la conmutación de control. Debemos dar en el punto buscado $(0, 0)$ $(x_1(T) = x_2(T) = 0)$ realizando no más de una conmutación y moviéndonos por la trayectoria fásica en la dirección permitida. El total de condiciones iniciales correspondientes a los controles $u^+(\cdot)$ y $u^-(\cdot)$ pueden ser descritas por las desigualdades $\xi_2 > \varphi(\xi_1)$ para $u^+(\cdot)$, y $\xi_2 < \varphi(\xi_1)$ para $u^-(\cdot)$ (fig. 3.) Dichas conmutaciones se efectúan en la curva $\xi_2 = \varphi(\xi_1)$. Además es fácil comprender que para cada condición inicial hay una sola curva fásica que conduce al punto $(0, 0)$. Como $|\dot x_2| = 1$, siempre se encuentra en la trayectoria óptima, resulta que $x_2 = |t| + C$ y que, por lo tanto, el tiempo de movimiento $\hat T = \mathrm{Var}\, x_2$.

Mostremos que la trayectoria óptima que comienza en el punto (ξ_1, ξ_2) proporciona la solución al problema. Supongamos que a esa trayectoria le corresponde el control $\hat u(\cdot)$ (para simplificar la tarea supongamos también que $u^-(\cdot)$), la función $\hat x(\cdot)$ y el tiempo $\hat T$. A su vez, admitamos que existe algún otro proceso controlado, por ejemplo, $(x(\cdot), u(\cdot), T)$, $T \leqslant \hat T$. Definamos complementariamente la función $x(\cdot)$ con un cero en el segmento $[T, \hat T]$.

En virtud de las condiciones en el extremo izquierdo, las funciones $\hat x(\cdot)$ y $x(\cdot)$ en el punto τ pueden ser representadas de la forma siguiente:

$$x(\tau) = \int_0^\tau (\tau - s)\ddot x(s)\,ds + \xi_2\tau + \xi_1.$$

Dado que $\ddot{\hat x}(s) = 1 \geqslant \ddot x(s)$ $\forall s \in [0, \tau]$, entonces

$$\hat x(\tau) - x(\tau) = \int_0^\tau (\tau - s)(1 - \ddot x(s))\,ds \geqslant 0,$$

además, la igualdad aquí sólo es posible cuando $\ddot{x}(s) \equiv 1$ en todos los puntos de continuidad, pero en este caso resulta
que $x(t) = \hat{x}(t)\ \forall t \in [0, \tau]$.

Análogamente, tomando en consideración las condiciones en el extremo derecho, la función $\hat{x}(\cdot)$ y $x(\cdot)$ en el punto τ puede ser representada de la forma siguiente:

$$x(\tau) = \int_{\tau}^{\hat{T}} (s - \tau)\, \ddot{x}(s)\, ds.$$

Como $\ddot{x}(s) \geqslant -1 = \ddot{\hat{x}}(s)\ \forall s \in (\tau, \hat{T})$, entonces

$$\hat{x}(\tau - x(\tau) = \int_{\tau}^{\hat{T}} (s - \tau)(-1 - \ddot{x}(s))\, ds \leqslant 0,$$

además, también aquí la igualdad sólo es posible cuando $\ddot{x}(s) \equiv -1$ y $x(t) = \hat{x}(t)\ \forall t \in [\tau, \hat{T}]$.

Así pues, tenemos $\hat{x}(\tau) = x(\tau)$ y, por consiguiente, $\hat{x}(t) \equiv (t)\, x(t)\ \forall t \in [0, \hat{T}]$. De aquí $T = \hat{T}$.