Критерий неотрицательности вырожденной квадратичной формы с двумерным управлением

А.В. Дмитрук

Как известно, квадратичные условия слабого минимума в задачах классического вариационного исчисления (а также в некоторых задачах оптимального управления) приводят к изучению знакоопределенности интегрального квадратичного функционала, и при этом изучении основным предположением является усиленное условие Лежандра.

Случай, когда условие Лежандра вырождается хотя бы в одной точке, изучен гораздо меньше (отметим здесь работы [5, 8, 3]). Однако он представляет интерес как с точки зрения теории самих квадратичных форм, так и потому, что он вполне может реализоваться в конкретных вариационных задачах. Если ограничиться вырождением условия Лежандра в одной точке, то первый интересный и нетривиальный функционал будет иметь следующий вид:

$$J = \int_0^1 \Big(t^2 (u,u) - 2bt (Px,u) + (Dx,x) \Big)\, dt, \tag{1}$$

$$\text{где} \qquad \dot x = u, \qquad x(1) = 0. \tag{2}$$

Здесь x и u двумерны, $P-$ матрица поворота на 90^o, $b \in \mathbb{R}-$ произвольный параметр, $D-$ постоянная симметричная матрица. Функцию $u(t)$ будем пока для простоты считать принадлежащей $L_\infty[0,1]$, т.е. $x(t)$ липшицевой. (Выбор знака минус в среднем члене (1) обусловлен удобством дальнейших формул.)

Вид этого функционала получается из тех соображений, что если для некоторого функционала с членом $(R(t)u,u)$ выполнено условие Лежандра: $R(t) \ge 0$, но оно вырождается в одной точке t_0, то типичное вырождение имеет вид $R = (t-t_0)^2$. Коэффициент при (Px,u) также должен вырождаться в t_0 (иначе этот член будет главным, и J заведомо будет иметь отрицательные значения), причем это вырождение должно иметь первый порядок (при большем вырождении смешанный член проигрывает обоим крайним по порядку величины). Естественно начать изучение этого функционала на вариациях $x(t)$, сосредоточенных в окрестности данной точки, т.е. при условиях $x(t_0 - \varepsilon) = x(t_0 + \varepsilon) = 0$, а тогда, ограничившись в силу симметрии рассмотрением полуинтервала $[t_0, t_0 + \varepsilon]$, приходим к виду (1), (2).

Обратим внимание, что функционал (1) обладает свойством автомодельности (или, если угодно, "фрактальности"): его неотрицательность не зависит от отрезка

$[0,T]$, на котором он рассматривается (при условии, что на правом его конце $x(T) = 0$), поэтому мы положили $T = 1$.

Поставим вопрос: при каких параметрах b и D функционал J будет неотрицательным на указанном множестве функций? (Тогда, очевидно, он будет неотрицательным и на его естественном замыкании, описание которого будет дано ниже.) Непосредственно применять классическое условие Якоби мы не можем, т.к. не выполнено усиленное условие Лежандра (оно вырождается в точке $t = 0$).

1. Установим сначала, что левый конец $x(t)$ также можно считать нулевым.

Лемма 1. Свойство $J(x) \geq 0$ для всех липшицевых $x(t)$ с граничным условием $x(1) = 0$ эквивалентно свойству $J(x) \geq 0$ для всех липшицевых $x(t)$ с граничными условиями $x(0) = x(1) = 0$.

Доказательство. Достаточно доказать, что если $J(\hat{x}) < 0$ для некоторой липшицевой $\hat{x}$ с условием $\hat{x}(1) = 0$, то найдется липшицевая x с условием $x(0) = x(1) = 0$, для которой также $J(x) < 0$.

Не ограничивая общности, можно считать, что $\hat{x} = const$, т.е. $\hat{u} = 0$ на отрезке $[0, \Delta]$ при некотором $\Delta > 0$ (т.к. множество таких u всюду плотно в $L_\infty[0,1]$ относительно любой интегральной метрики). Возьмем любое $\varepsilon < \Delta$ и построим функцию $x(t)$, которая на $[0,\varepsilon]$ линейно растет от 0 до $\hat{x}(\varepsilon)$, а затем совпадает с $\hat{x}(t)$. Таким образом, построенные x, u отличаются от исходных $\hat{x}, \hat{u}$ лишь на отрезке $[0,\varepsilon]$, на котором $|u(t)| \simeq 1/\varepsilon$ (величина порядка $1/\varepsilon$), и следовательно, $|tu| \leq \mathrm{const}$, и также $t^2\,|u|^2 \leq \mathrm{const}$. Поэтому $|J(x) - J(\hat{x})| \leq \mathrm{const} \cdot \varepsilon$, и тогда при малом ε получим $J(x) < 0$, ч. т. д.

Эта лемма позволяет, в принципе, применять условия Якоби, т.к. теперь можно считать $x(0) = 0$, и тогда, согласно общей идее этих условий, можно двигать левый конец отрезка. На каждом отрезке $[\theta, 1]$ при $\theta > 0$ усиленное условие Лежандра выполнено, и поэтому можно искать точку t_*, сопряженную с точкой $t = 1$. Если такой точки нет на интервале (0,1), то для любого $\theta > 0$ функционал $J \geq 0$ на $[\theta, 1]$, а тогда по непрерывности $J \geq 0$ и на $[0, 1]$. (Эту же процедуру можно проделать и для свободного $x(0)$, не пользуясь леммой 1, но тогда при рассмотрении J на $[0,1]$ надо добавить к нему внеинтегральный член $(Dx(\theta), x(\theta))\cdot\theta$, соответствующий интегралу по $[0,\theta]$, и коэффициенты функционала перестанут быть постоянными.)

Однако мы будем действовать иначе. Взяв произвольную постоянную симметричную матрицу S, добавим под знак интеграла в (1) выражение

$$-\frac{d}{dt}\,[t(Sx,x)] = \; -2t(Sx,u) - (Sx,x),$$

которое, очевидно, не меняет значение функционала (ибо интеграл от него есть $(t(Sx,x))|_0^T = 0$). Тогда подинтегральное выражение примет вид

$$[tu - (S + bP)x]^2 + (Mx,x),$$

$$\text{где} \quad M = D - S - S^2 - b^2E + b(PS - SP),$$

$E-$ единичная матрица. Если при некоторой S получим $M \geq 0,$ то неотрицательность J очевидна. Если же такой S подобрать не удастся, то покажем, что тогда $J(x) < 0$ для некоторого $x.$

2. Прежде, чем проводить эту программу, заметим, что функционал (1) имеет еще две интересные переформулировки. При замене $t = e^{-\tau},\ \frac{dt}{t} = -d\tau,$ и соответственно $tu = t\frac{dx}{dt} = -\frac{dx}{d\tau} = -w(\tau),$ он превращается в функционал

$$J = \int_0^\infty e^{-\tau} \left[w^2 + 2b(Px, w) + (Dx, x)\right] d\tau, \tag{3}$$

$$\frac{dx}{d\tau} = w, \qquad x(0) = 0.$$

Подобные функционалы типичны для моделей математической экономики; коэффициент $e^{-\tau}$ называется дисконтирующим множителем. Знак минус в среднем члене в формуле (1) был выбран с тем расчетом, чтобы функционал (3) имел бы "канонический"вид со знаком плюс.

Если теперь положить $e^{-\tau/2}x = z,\ \ e^{-\tau/2}w = v,$ то получим линейную систему с постоянными коэффициентами

$$\frac{dz}{d\tau} = -\frac{1}{2}\,z + v, \qquad z(0) = 0, \tag{4}$$

и функционал, также имеющий постоянные коэффициенты:

$$J = \int_0^\infty \left[v^2 + 2b(Pz, v) + (Dz, z)\right] d\tau. \tag{5}$$

Чтобы не заботиться о сходимости интеграла, будем изучать этот функционал на всех финитных $z(t),$ т.е. таких, что $z(t) = 0$ для всех достаточно больших $t;$ при этом в силу уравнения (4) и $v(t)$ также будет финитным. (Для функционала (1) это соответствует рассмотрению лишь тех $x,$ для которых $x(t) = 0$ на $[0, \varepsilon]$ при некотором $\varepsilon > 0.$)

Дифференциальную связь (4) можно упростить, положив $\frac{dz}{d\tau} = u = -\frac{1}{2}\,z + v,$ выразив отсюда $v = u + \frac{1}{2}\,z$ и подставив в (5). При этом $(Pz, z) = 0,$ а члены типа $\int 2(z, \dot z)\, dt = (z, z)\big|_0^\infty = 0$ пропадут. Таким образом, заменив обозначения τ, z на обычные $t, x,$ получим

$$\dot x = u, \qquad x(0) = 0, \tag{6}$$

$$J = \int_0^\infty \left[u^2 + 2b(Px, u) + (Qx, x)\right] dt, \tag{7}$$

где $Q = D + \frac{1}{4}E.$ Именно в этой форме мы будем далее изучать наш функционал. Установим сначала простое свойство, которое верно для функционала (6, 7) в пространстве $\mathbb{R}^n$ любой размерности.

Лемма 2. Если $J \geq 0,$ то $Q \geq 0$ (условие типа Лежандра).

Доказательство проведем от противного. Допустим, что существует такой $h \in \mathbb{R}^n,$ что $(Qh, h) < 0.$ Положим $x(t) = h,\ u(t) = 0$ на отрезке $[1, T],$ а на отрезках

[0,1] и [T, T+1] пусть x линейно меняется от 0 до h и от h до 0 соответственно. Т.к. интеграл по отрезку [1,T] есть отрицательная величина $(T-1)(Qh,h)$ порядка T, а интегралы по отрезкам [0,1] и [T, T+1] конечны, то при больших T получим $J(x)<0$, противоречие.

3. Следуя вышеуказанной идее, возьмем произвольную симметричную матрицу S и добавим под интеграл выражение $\frac{d}{dt}(Sx,x) = 2(Sx,u)$ так, чтобы выделить полный квадрат от членов, содержащих u. Тогда получим

$$J = \int_0^\infty \Big([u + (S+bP)x]^2 + (Mx,x) \Big)\, dt, \tag{8}$$

$$\text{где} \qquad (Mx,x) = (Qx,x) - (Sx+bPx)^2,$$

$$\text{т.е.} \qquad M = Q - (S+bP)^*(S+bP) = Q - S^2 + b(PS-SP) - b^2E. \tag{9}$$

Если при этом получилось $M \geq 0$, то очевидно $J \geq 0$ на всех финитных функциях, удовлетворяющих (6).

Спрашивается, при каких b и Q можно, подбирая подходящую матрицу S, получить $M \geq 0$?

Без нарушения общности можно считать, что матрица Q диагональна (т.к. при повороте двумерного вектора x и соответствующего вектора u квадратичные формы (u,u) и (Px,u) не изменяются), т.е.

$$Q = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}, \qquad \text{где согласно лемме 2} \qquad q_1 \geq 0, \quad q_2 \geq 0.$$

Будем искать S в виде

$$S = \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}, \qquad \text{где } c - \text{неизвестный пока параметр.}$$

(Ниже мы увидим, что рассмотрение таких S для наших целей достаточно.) Тогда, выполняя простые калькуляции, получим

$$M = Q - c^2E - b^2E + 2bc \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} q_1 - (b+c)^2 & 0 \\ 0 & q_2 - (b-c)^2 \end{pmatrix}, \tag{10}$$

и вопрос сводится к следующему: при каких b, q_1, q_2 найдется такое c, что

$$q_1 - (b+c)^2 \geq 0 \qquad \text{и} \qquad q_2 - (b-c)^2 \geq 0 \ ?$$

Последнее эквивалентно тому, что при некотором c

$$\sqrt{q_1} \geq |b+c| \qquad \text{и} \qquad \sqrt{q_2} \geq |b-c|. \tag{11}$$

Нетрудно сообразить, что такое c найдется в том и только в том случае, когда

$$\frac{1}{2}\,(\sqrt{q_1}+\sqrt{q_2}) \;\geq\; |b|. \tag{12}$$

Итак, мы установили, что если выполнено (12), то найдется симметричная матрица S (указанного вида) такая, что соответствующая $M \geq 0$, и следовательно, $J \geq 0$.

Что будет, когда (12) не выполнено? Требуемой матрицы S указанного вида (с нулевой диагональю) нет, но означает ли это нарушение свойства $J \geq 0$? Мы покажем, что это действительно так, и в этом и состоит ключевой пункт предлагаемого здесь подхода.

Не ограничивая общности будем для удобства считать $b \geq 0$ (иначе сделаем замену $x_1 \leftrightarrow x_2,\; u_1 \leftrightarrow u_2,$ при которой (Px,u) перейдет в $-(Px,u),$ а остальные члены в J не изменятся.)

Итак, пусть

$$\frac{1}{2}\,(\sqrt{q_1}+\sqrt{q_2}) \;<\; b.$$

В этом случае очевидно найдется такое $c,$ что

$$\sqrt{q_1} < b + c, \qquad \sqrt{q_2} < b - c \tag{13}$$

(например, можно взять $c = \frac{1}{2}\,(\sqrt{q_1}-\sqrt{q_2})\,),$ т.е.

$$q_1^2 < (b+c)^2, \qquad q_2^2 < (b-c)^2,$$

— оба элемента матрицы M отрицательны, следовательно, $M < 0$ (отрицательно определена). Тогда при некотором $\delta > 0$ имеем $(Mx,x) \leq -\delta |x|^2 \quad \forall\, x,$ т.е. второй подинтегральный член в (8) отрицательно определен.

Попробуем теперь найти такую допустимую пару $x,u,$ на которой первый подинтегральный член в (8) был бы нулевым, т.е. положим $u + (S+bP)x = 0.$ Тогда получаем уравнение

$$\dot x = -(S+bP)x. \tag{14}$$

Лемма 3. При выполнении (13) уравнение (14) имеет периодическое решение вида $x = f\sin\omega t + h\cos\omega t$ при некотором $\omega \neq 0$ и неколинеарных векторах $f,h \in \mathbb{R}^2$.

Доказательство. Достаточно показать, что матрица $R = (S+bP)$ имеет чисто мнимое собственное значение $\lambda = i\omega \neq 0.$ Так как

$$R = \begin{pmatrix} 0 & c-b \\ c+b & 0 \end{pmatrix},$$

то уравнение для собственных значений имеет вид $\lambda^2 - (c^2-b^2) = 0.$ Если $c^2 - b^2 < 0,$ т.е. $|c| < b,$ то это уравнение имеет два чисто мнимых корня $\lambda = \pm i\omega,\; \omega =$

$\sqrt{b^2-c^2} > 0$. Покажем, что у нас реализуется именно этот случай. Из неравенств (13) следует, что

$$-c < b-\sqrt{q_1} \le b, \qquad c < b-\sqrt{q_2} \le b,$$

т.е. $\pm c < b$, а это и означает, что $|c| < b$. Лемма доказана.

Так как векторы f, h линейно независимы, то $x(t) = f\sin\omega t + h\cos\omega t$ описывает эллипс в $\mathbb{R}^2$, и можно считать, что на нем $|x| \ge 1$. Рассмотрим теперь указанное решение $x(t)$ на большом отрезке времени $[1,T]$ (подобно тому, как ранее мы рассматривали $x = \mathrm{const}$). На нем первый член (квадрат) в выражении (8) по определению равен нулю, а т.к. $|x| \ge 1$, то $(Mx,x) \le -\delta$, поэтому интеграл по отрезку $[1,T]$ есть отрицательная величина $\le -\delta(T-1)$ порядка T. На отрезках [0,1] и $[T, T+1]$, как и раньше, сведем x к нулевым концевым значениям в точках 0 и $T+1$; интегралы по этим отрезкам в силу ограниченности $x(t)$ опять дадут лишь конечный вклад в функционал. Поэтому на всем отрезке $[0, T+1]$ при большом T получим функцию $\hat{x}(t)$, для которой $J(\hat{x}) \le -\delta T/2 < 0$.

Замечание. Подчеркнем еще раз: важный момент здесь состоит в том, что найденное циклическое решение уравнения (14) мы можем "крутить"сколь угодно долго, набирая тем самым сколь угодно большой отрицательный интеграл от (Mx,x) и сохраняя ограниченное значение x. При этом первый член в (8) все время остается равным нулю в силу (14).

Итак, мы показали, что если неравенство (11) не выполнено, то найдется финитная функция $\hat{x}(t)$, для которой $J(\hat{x}) < 0$. Тем самым, с учетом предыдущего, мы установили следующий факт.

Теорема 1. *Функционал (7) неотрицателен на всех финитных функциях, удовлетворяющих уравнению (6), тогда и только тогда, когда собственные значения матрицы Q неотрицательны и удовлетворяют неравенству (12).*

Замечание. Возводя (12) в квадрат, получаем эквивалентное неравенство $(q_1+q_2) + 2\sqrt{q_1 q_2} \ge 4b^2$, которое можно записать в терминах исходной матрицы Q, не приводя ее к диагональному виду:

$$\mathrm{Tr}\, Q \;+\; 2\sqrt{det\, Q} \;\ge\; 4b^2. \tag{15}$$

4. Частные случаи. а) Пусть $b=0$, $Q=0$. Тогда в (12) выполнено равенство, и по теореме 1 имеем $J = \int_0^\infty u^2\,dt \ge 0$ для любого $x(t)$, такого что $\dot{x} = u$, $x(0)=0$. Этот результат, конечно, очевиден и без теоремы 1.

Посмотрим, что означают полученные условия для функционалов (1) и (3). Так как $Q = D + \frac{1}{4}E$, то неравенство $Q \ge 0$ означает $D + \frac{1}{4}E \ge 0$, а неравенство (12) — что

$$\frac{1}{2}\left(\sqrt{d_1+\frac{1}{4}} + \sqrt{d_2+\frac{1}{4}}\right) \;\ge\; |b|. \tag{16}$$

b) Рассмотрим случай $b = 0$, $D = -\frac{1}{4}E$ (т.е. $Q = 0$) . Тогда функционал (3) имеет вид (мы опять пишем t вместо τ):

$$J = \int_0^\infty e^{-t}(u^2 - \frac{1}{4}x^2)\, dt \;\geq\; 0 \tag{17}$$

на всех финитных $x(t)$ таких, что

$$\dot{x} = u, \qquad x(0) = 0. \tag{18}$$

Другими словами, для таких функций выполняется неравенство

$$\int_0^\infty e^{-t}x^2\, dt \;\leq\; 4\int_0^\infty e^{-t}u^2\, dt. \tag{19}$$

Если ввести гильбертово пространство $H = L_2[0,\infty)$ с весом e^{-t}, то (в силу того, что финитные функции всюду плотны в нем) из (19) следует, что интегральный оператор $u \mapsto x$, заданный по формуле (18), есть линейный ограниченный оператор $H \to H$, и его норма не превосходит $\sqrt{4} = 2$. В действительности его норма равна 2, так как константа 4 в неравенстве (19) точная.

c) Для функционала (5) это же свойство означает, что для любой финитной функции z, удовлетворяющей (4),

$$J = \int_0^\infty (v^2 - \frac{1}{4}z^2)\, dt \;\geq\; 0,$$

$$\text{т.е.} \qquad \int_0^\infty z^2\, dt \;\leq\; 4\int_0^\infty v^2\, dt. \tag{20}$$

Отсюда следует, что для любой функции $v \in L_2[0,\infty)$, функция $z(t)$, удовлетворяющая уравнению

$$\dot{z} = -\frac{1}{2}\, z + v, \qquad z(0) = 0, \tag{21}$$

также принадлежит $L_2[0,\infty)$, причем норма оператора $v \mapsto z$ не превосходит (а в действительности равна) 2.

Отметим, что тот же факт справедлив и для уравнения

$$\dot{z} = -kz + v, \qquad z(0) = 0, \tag{22}$$

при любом $k > 0$. (Простым масштабированием оно сводится к (21).) А именно, справедлива

Лемма 4. *Для любой функции $v \in L_2[0,\infty)$, функция $z(t)$, являющаяся решением уравнения (22), также принадлежит $L_2[0,\infty)$, причем норма оператора $v \mapsto z$ равна $1/k$.*

Замечание. Указанный оператор, хотя и является интегральным, и поэтому вполне непрерывным в пространстве $L_2[0,T]$ при любом конечном T, не является вполне непрерывным в пространстве $L_2[0,\infty)$. В частности, он имеет чисто непрерывный спектр. (На комплексной плоскости это круг, диаметр которого есть отрезок $[0, 1/k]$ действительной оси.)

d) Для функционала (1) при $D = -\frac{1}{4}E$ и $b = 0$ получаем неравенство

$$J = \int_0^1 \left(t^2 u^2 - \frac{1}{4} x^2 \right) dt \ \geq \ 0,$$

$$\text{т.е.} \qquad \int_0^1 x^2 \, dt \ \leq \ 4 \int_0^1 \ t^2 \, u^2 \, dt \tag{23}$$

для всех $x, u,$ связанных уравнением (2) и равных нулю в некоторой окрестности $t = 0$. Тогда это же неравенство верно и для всех $u(t),$ для которых сходится интеграл в правой части (23) (т.е. для всех $u(t)$ из пространства $L_2[0,1]$ с весом t^2); при этом интеграл в левой части также будет сходиться в силу оценки (23). Отметим, что во всех трех неравенствах (19, 20, 23) размерность x может быть произвольной, т.к. фактически эти неравенства одномерны.

Неравенство (23) и соответствующие ему неравенства (20) и (19) есть известное классическое неравенство Г.Харди [4, §9.8], поэтому неравенство $J \geq 0$ при выполнении (12) и (16) можно трактовать как его двумерное обобщение.

e) Итак, функционалы (1, 3, 5, 7) имеют смысл не только для финитных x, u (что мы предполагали вначале для простоты), а для любых пар x, u из пространства L_2 с соответствующим весом. Имеет место следующее простое

Предложение 1. В функционале (3) пара $(x, w) \in L_2\,[0,\infty)$ с весом $e^{-\tau}$ $\Longleftrightarrow$ в функционале (5) пара $(z, v) \in L_2\,[0,\infty)$ "обычному" $\Longleftrightarrow$ в функционале (7) пара $(x, u) \in L_2\,[0,\infty)$ $\Longleftrightarrow$ в функционале (1) пара $(x,\ tu) \in L_2\,[0,1]$.

Неотрицательность функционала J на множестве финитных пар эквивалентна его неотрицательности на множестве пар из соответствующего пространства L_2. (Для пространства $L_2\,[0,1]$ считаем финитной пару, у которой $x = u = 0$ на $[0,\varepsilon]$ при некотором $\varepsilon > 0$.)

Из оценок (19, 20, 23) для функционалов (1, 3, 5) вытекает, что принадлежность управления соответствующему пространству L_2 обеспечивает принадлежность и фазовой компоненты этому же пространству L_2. Для функционала (7) это не так: здесь соответствующий случай $b = 0,\ Q = 0$ приводит, как уже отмечалось, к тривиальному неравенству $\int u^2 \, dt \ \geq \ 0,$ которое не связывает фазовую компоненту с управлением.

Укажем, как соотносятся представления в виде "суммы квадратов"для функционалов (1, 3, 5, 7). Если для функционала (1) мы получили

$$J = \int_0^1 ((tu - Rx)^2 + (Mx, x))\, dt, \qquad\qquad \frac{dx}{dt} = u,$$

то для (3) после замены $dt = e^{-\tau} d\tau,\ \ tu = -w$ получаем

$$J = \int_0^\infty e^{-\tau} \left[(w + Rx)^2 + (Mx, x) \right] \, d\tau, \qquad\qquad \frac{dx}{d\tau} = w,$$

для функционала (5) имеем $e^{-\tau/2}x = z,\;\; e^{-\tau/2}w = v,$

$$J = \int_0^\infty \Big[(v + Rz)^2 + (Mz, z)\Big]\, d\tau, \qquad\qquad \frac{dz}{d\tau} = -\frac{1}{2}\, z + v,$$

и тогда для (7) имеем $u = -\frac{1}{2}\, z + v,$ т.е. $v = \frac{1}{2}\, z + u,$

$$J = \int_0^\infty \Big[(u + (R + \frac{1}{2})z)^2 + (Mz, z)\Big]\, d\tau, \qquad\qquad \frac{dz}{d\tau} = u.$$

5. Обратим внимание на одну интересную деталь.

Лемма 5. *Пусть для функционала (7) неравенство (12) или эквивалентное ему (15) выполнено со знаком равенства. Тогда функционал J положителен, т.е. $J(x) > 0 \quad \forall\, x \neq 0.$*

Действительно, в случае равенства в (12) в формуле (11) также выполняются равенства, и матрица $M = 0.$ Поэтому функционал J сводится к первому члену в (8), и если $J(x) = 0,$ то $u + (S + bP)x = 0,$ т.е. $\dot x = -(S + bP)x.$ Так как при этом $x(0) = 0,$ то получаем $x \equiv u \equiv 0,$ ч.т.д.

Таким образом, если (12) выполнено в виде равенства, то, с одной стороны, всегда $J(x) > 0,$ а с другой – уменьшить какое-то q_i или увеличить $|b|$ нельзя, т.к. тогда неравенство (12) нарушится, и по той же теореме 1 получим $J(x) < 0$ при некотором x .

Это обстоятельство на первый взгляд противоречит кажущейся лежандровости функционала J (особенно в формуле (5) — при u^2 стоит единичная матрица, т.е. равномерно по t выполнено усиленное условие Лежандра!), ибо известно, что если лежандровый функционал положителен, то он положительно определен, и поэтому любой близкий к нему функционал также будет положительно определенным. Однако никакого противоречия на самом деле нет, т.к. наш J в действительности не является лежандровым: для функционалов на $[0, \infty)$ положительности коэффициента при u^2 еще не достаточно для лежандровости, т.к. оставшиеся члены не являются слабо непрерывными. (Это соответствует тому, что упомянутый в лемме 4 оператор $u \mapsto x$ не является вполне непрерывным при рассмотрении его на бесконечном интервале.)

Из леммы 5 вытекает, что константа 4 в неравенствах (19, 20, 23) точная, но она не достигается. Т.е., с одной стороны, (а) ее нельзя уменьшить, а с другой, (б) для любой ненулевой функции $x(t)$ эти неравенства строгие, т.е. выполнены с некоторой $C(x) < 4.$ Это довольно любопытная особенность неравенств для функций и их производных.

Еще одно следствие леммы 5 — это возможность построения примера задачи на $[0, \infty),$ в которой не существует решения.

Пример 1. Рассмотрим задачу $J(x,u) \to \min$, где J задан, например, формулами (6, 7) с коэффициентами, удовлетворяющими (12) в виде равенства (напр., $b = q_1 = q_2 = 0$), при ограничении

$$\int_0^\infty |x|^2\, dt \;=\; 1. \tag{24}$$

Покажем, что в этой задаче минимум не достигается. Для этого достаточно показать, что $\inf J = 0$ (а тогда он заведомо не достигается по лемме 5).

Действительно, если при заданном ограничении $\inf J = a > 0$, то в силу однородности для любого $x \in L_2\,[0,\infty)$ будет $J(x) \ge a \int |x|^2 dt$, т.е.

$$\tilde J(x) = \; J(x) - a \int |x|^2 dt \;\ge 0. \tag{25}$$

Но у функционала $\tilde J$ матрица $\tilde Q = Q - aE$, и неравенство (12) для нее нарушается (поскольку для Q оно выполнено как равенство); следовательно, по теореме 1 не может для всех x выполняться (25), противоречие.

Обратим внимание, что предложенная задача удовлетворяет стандартному требованию выпуклости по u. Можно, правда, возразить, что она не удовлетворяет требованию компактности по u. Рассмотрим тогда

Пример 1'. — та же задача с дополнительным ограничением $|u| \le 1$. Покажем, что в ней по прежнему $\inf J = 0$.

Пусть для определенности $b \ge 0$. Если в (12) выполнено равенство, то существует единственное c, для которого выполняется (11), причем оно превращается в равенство. В этом случае $\sqrt{q_1} = b + c$, $\sqrt{q_2} = b - c$, матрица $M = 0$, а функционал имеет вид

$$J = \int_0^\infty (u + Rx)^2\, dt, \tag{26}$$

$$\text{где} \qquad R = \begin{pmatrix} 0 & -(b-c) \\ (b+c) & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\sqrt{q_1} \\ \sqrt{q_2} & 0 \end{pmatrix}. \tag{27}$$

Так как матрица R имеет характеристическое уравнение $\lambda^2 + \sqrt{q_1 q_2} = 0$, то она всегда имеет собственное значение $\lambda = i\omega$ с нулевой вещественной частью, и поэтому уравнение (14) $\dot x = -Rx$ всегда имеет решение вида $x = \xi\, e^{i\omega t}$ при некотором действительном ω и комплексном $\xi \ne 0$.

Как и в лемме 3, для каждого N возьмем функцию $x_N(t)$, совпадающую с этим решением на отрезке $[1, N]$, зануляющуюся при $t = 0$ и $t > N+1$ и линейную на отрезках $[0,1]$ и $[N, N+1]$. При этом $|x_N(t)| \le \text{const}$, и для $u_N = \dot x_N$ также справедливо $|u_N(t)| \le \text{const}$. Следовательно,

$$J(x_N) = \int_0^1 (u + Rx)^2\, dt \;+ \int_N^{N+1} (u + Rx)^2 dt \;\le\; \text{const}$$

$$\text{и} \qquad \int_0^\infty |x_N|^2 dt \;=\; \beta_N^2 \to \infty.$$

(Здесь β_N^2 есть величина порядка N).
Тогда $\hat{x}_N = x_N/\beta_N$ удовлетворяет ограничению (24), для него $|\hat{u}_N| \le \frac{\text{const}}{\beta_N} \to 0$, поэтому при больших N пара $(\hat{x}_N, \hat{u}_N)$ удовлетворяет всем ограничениям, и на ней

$$J(\hat{x}_N) \le \frac{\text{const}}{\beta_N^2} \to 0.$$

Отсюда $\inf J = 0$, ч.т.д.

В частном случае, когда $b = q_1 = q_2 = 0$ (фактически это одномерный случай), получаем, что в задаче

$$J = \int_0^\infty u^2 dt \to \min, \qquad \dot{x} = u, \quad x(0) = 0,$$

$$\int_0^\infty x^2 dt = 1,$$

а также в этой задаче с дополнительным ограничением $|u| \le 1$ минимум не достигается.

Аналогичные соображения показывают, что нет решения и в следующем примере "экономического"типа.

Пример 2. Рассмотрим задачу

$$J = \int_0^\infty e^{-\tau} \left[w^2 + 2b(Px, w) + (Dx, x) \right] d\tau \to \min,$$

$$\frac{dx}{d\tau} = w, \qquad x(0) = 0,$$

$$\int_0^\infty e^{-\tau}(x, x)\, d\tau = 1, \tag{28}$$

где b и D удовлетворяют условию (16) в виде равенства (напр., $b = 0$, $d_1 = d_2 = -1/4$).

Здесь, как и раньше, $\inf J = 0$, но он не достигается. Действительно, в этом случае соответствующий функционал вида (7), как мы знаем, представим в виде (26) с матрицей R вида (27), имеющей собственное значение $\lambda = i\omega$, а тогда, согласно п. 4, наш функционал представим в виде

$$J = \int_0^\infty e^{-\tau} (w + R'x)^2 \, d\tau,$$

где $R' = R - \frac{1}{2} E$ имеет собственное значение $\lambda = -\frac{1}{2} + i\omega$. Тогда уравнение $\dot{x} = -R'x$ имеет решение $x(\tau) = \xi\, e^{(\frac{1}{2} + i\omega)\tau}$, и для соответствующего $x_N(\tau)$ (совпадающего с $x(\tau)$ на отрезке $[1, N]$ и равного нулю вне $(0, N{+}1)$) будем по-прежнему иметь

$$\int_0^\infty e^{-\tau} |x_N|^2 \, d\tau \;=\; \beta_n \to \infty.$$

При этом для $u_N(\tau) = \dot{x}_N(\tau)$ всюду выполнена оценка $|u_N(\tau)| \le \text{const} \cdot e^{\tau/2}$, так что по-прежнему $J(x_N) \le \text{const}$. Далее, переходя, как и раньше, к $\hat{x}_N = x_N/\beta_N$, получим $J(\hat{x}_N) \to 0$, ч.т.д.

Отметим однако, что при переходе к задаче $2'$ с ограничением $|u| \le c$ аналогии с задачей $1'$ уже нет. Дело в том, что в примере 2 "зануляющее"управление $u = \xi\, e^{(\frac{1}{2}+i\omega)\tau}(\frac{1}{2}+i\omega)$ имеет норму $||u||_\infty$ порядка $e^{\frac{1}{2}N}$ (она достигается при $\tau = N$), тогда как левая часть в (28) есть величина порядка $N,$ поэтому для нормированной последовательности $\hat{x}_N = x_N/\sqrt{N},$ удовлетворяющей (28), получаем $||\hat{u}_N||_\infty \simeq e^{N/2}/\sqrt{N} \to \infty,$ что означает нарушение ограничения $||u||_\infty \le c$ при любом $c.$

6. Связь с теорией сопряженных точек. Исследуем теперь вопрос о неотрицательности функционала (6, 7) с помощью классического условия Якоби о сопряженных точках. Неотрицательность J на множестве всех финитных x, u (которая эквивалентна его неотрицательности на всех $x, u \in L_2\,[0,\infty)$) очевидно эквивалентна тому, что $\forall T > 0 \quad J(x,u) \ge 0$ для всех $x, u,$ сосредоточенных на отрезке $[0, T].$

Таким образом, нам надо, чтобы для любого фиксированного $T > 0$ функционал

$$J_T = \int_0^T \Big[u^2 + 2b(Px, u) + (Qx, x)\Big]\; dt \ge 0 \tag{29}$$

на всех x, u таких, что $u \in L_2\,[0, T],$

$$\dot{x} = u, \qquad x(0) = 0, \quad x(T) = 0. \tag{30}$$

Мы имеем стандартный квадратичный функционал классического вариационного исчисления, удовлетворяющий усиленному условию Лежандра. Условие Якоби для него гласит, что $J_T \ge 0$ тогда и только тогда, когда интервал $(0, T)$ не содержит точки, сопряженной с $t = 0.$ Поскольку T у нас произвольно, то вообще на всей полупрямой $(0, \infty)$ не должно быть точек, сопряженных с $t = 0.$ Мы хотим выяснить, при каких b и Q реализуется этот случай.

Для нахождения сопряженной точки выпишем уравнение Эйлера–Якоби:

$$\frac{d}{dt}\,(L_u) \;=\; L_x\,,$$

где L есть подинтегральное выражение в (29). Таким образом, имеем уравнение (второго порядка, с постоянными коэффициентами)

$$\ddot{x} \;=\; -(2b)\,P\dot{x} \,+\, Qx, \qquad\quad x(0) = 0, \tag{31}$$

а сопряженная точка T_* определяется равенством $x(T_*) = 0.$ (Точнее, интересующая нас "первая"сопряженная точка — это наименьшее $T_* > 0,$ для которого существует нетривиальное решение уравнения (31), обращающееся в нуль в этой точке.)

Неотрицательность функционала (29) при любом T эквивалентна тому, что любое нетривиальное решение уравнения (31) не обращается в нуль нигде на $(0, \infty).$ Поэтому теорема 1 эквивалентна следующему (априори неочевидному) утверждению о качественном поведении решений уравнения (31).

Теорема 2. Для того, чтобы любое нетривиальное решение уравнения (31) не обращалось в нуль на полупрямой $(0, \infty)$, необходимо и достаточно, чтобы собственные значения матрицы Q были неотрицательны и удовлетворяли неравенству (12).

В случае, когда $Q = qE$ есть скалярная матрица, эту теорему нетрудно доказать и непосредственно. Неравенство (12) в этом случае приобретает вид

$$\sqrt{q} \geq b, \qquad \text{т.е.} \qquad q \geq b^2. \tag{32}$$

Будем считать вектор $x \in \mathbb{R}^2$ представлением комплексного числа $z \in \mathbf{C}$. Тогда матрица P соответствует умножению на i, и уравнение (31) приобретает вид

$$\ddot{z} \;=\; -(2bi)\,\dot{z} \;+\; qz, \qquad\qquad z(0) = 0. \tag{33}$$

Соответствующее характеристическое уравнение $\lambda^2 + (2bi)\,\lambda - q = 0$ имеет корни

$$\lambda = \; -bi \,\pm \sqrt{-b^2 + q}\;.$$

Под знаком корня как раз стоит величина из неравенства (32).
Рассмотрим все возможные случаи.

a) $q > b^2$, т.е. неравенство (32) выполнено строго. Тогда $\lambda = \pm a - bi$, где $a = \sqrt{q - b^2} > 0$, и при этом

$$z = e^{-ibt}\,(c_1 \operatorname{sh} at + c_2 \operatorname{ch} at).$$

В силу начального условия $z(0) = 0$ имеем $c_2 = 0$, поэтому $c_1 \neq 0$ (мы рассматриваем только нетривиальные решения!), и тогда $z(t) \neq 0$ для всех $t > 0$. Таким образом, в этом случае сопряженной точки нет.

b) $q = b^2$, т.е. (32) выполнено со знаком равенства. Тогда $\lambda = -bi$ есть характеристическое число кратности 2, и общее решение уравнения (33) будет иметь вид

$$z = e^{-ibt}\,(c_1 t + c_2).$$

С учетом начального условия имеем $c_2 = 0$, $c_1 \neq 0$, и тогда опять $z(t) \neq 0$ для всех $t > 0$, т.е. по-прежнему сопряженная точка отсутствует.

c) $0 < q < b^2$, т.е. (32) не выполнено. Тогда $\lambda = -ib \pm ia$, где $a = \sqrt{b^2 - q} < b$, т.е. $\lambda_1 = -i\mu_1$, $\lambda_2 = -i\mu_2$, где $0 < \mu_1 < \mu_2$. Здесь

$$z = c_1 e^{-i\mu_1 t} + c_2 e^{-i\mu_2 t},$$

и $c_1 + c_2 = 0$, т.е. $z = c\,(e^{-i\mu_1 t} - e^{-i\mu_2 t})$, $c \neq 0$. Равенство $z(T_*) = 0$ означает, что

$$e^{i(\mu_2 - \mu_1)T_*} = 1.$$

Наименьшее решение этого уравнения $T_* = 2\pi/(\mu_2 - \mu_1)$.
Таким образом, в этом случае сопряженная точка существует.

d) $0 = q < b^2$, (32) не выполнено. Тогда $\lambda_1 = 0$, $\lambda_2 = -2bi$, $z = c\,(-1 + e^{-2bi\,t})$, $c \neq 0$. Из равенства $z(T_*) = 0$ находим сопряженную точку $T_* = \pi/b$.

e) $q < 0 < b^2$, (32) не выполнено. Тогда $\lambda_1 = i\mu_1$, $\lambda_2 = -i\mu_2$, где $0 < \mu_1 < \mu_2$. Аналогично случаю c), здесь $z = c(e^{i\mu_1 t} - e^{-i\mu_2 t})$, $c \neq 0$, и сопряженная точка $T_* = 2\pi/(\mu_2 + \mu_1)$.

e) Наконец, рассмотрим тривиальный случай, когда $b = 0$. В этом случае $\lambda^2 = q$ (а функционал фактически становится одномерным). Если $q = 0$, т.е. (32) выполнено, то с учетом начального условия имеем $z = ct$, $\neq 0$, и тогда $z(t) > 0$ для всех $t > 0$.

Если $q > 0$, т.е. (32) также выполнено, то $z = c\,\mathrm{sh}\,\omega t$, $\omega = \sqrt{|q|}$, $\neq 0$, и опять $z(t) > 0$ для всех $t > 0$.

Если же $q < 0$, (32) не выполнено, то $z = c\,\sin\omega t$, $\omega = \sqrt{|q|}$, $\neq 0$, и условие $z(T_*) = 0$ дает сопряженную точку $T_* = \pi/\omega$.

Итак, мы рассмотрели все возможные случаи, и установили, что в каждом из них неравенство (32) правильно указывает на наличие или отсутствие сопряженной точки. Теорема 2 (а значит и эквивалентная ей теорема 1) доказана.

В случае различных собственных чисел $q_1 \neq q_2$ подобный способ доказательства теоремы 2 потребовал бы гораздо более сложных вычислений.

7. Связь с частотным критерием. Укажем также на еще один возможный способ исследования нашего функционала, а именно, на использование т.н. "частотного критерия"(см. напр. [1, 2]). Он, по сути, состоит в следующей простой процедуре (как ни странно, не упомянутой в указанных работах). Будем, как и выше, рассматривать функционал J вида (7) на решениях уравнения (6). Если $J(x) \geq 0$ на всех финитных x, то для любого фиксированного $T > 0$ очевидно имеем $J(x) \geq 0$ на всех функциях x, сосредоточенных на отрезке $[0, T]$. Любую такую функцию x и ее производную u можно разложить в ряд Фурье на этом отрезке. Если использовать комплексные обозначения, то

$$x = \mathrm{Re}\left(\xi_0 + \sum_{k=1}^{\infty} \xi_k\, e^{i\omega_k t}\right), \qquad u = \mathrm{Re}\left(\sum_{k=1}^{\infty} i\omega_k\, \xi_k\, e^{i\omega_k t}\right), \tag{34}$$

где $\xi_k -$ произвольные векторы из $\mathbf{C}^2$, ряд из квадратов которых сходится, а $\omega_k = \frac{T}{2\pi}k$. (Ясно, что достаточно рассматривать конечные суммы указанного вида, т.к. они плотны в $L_2[0, T]$.)

Подставим (34) в функционал (7). Заметим, что при этом произведения разных гармоник дадут функцию вида $\xi'\, e^{i\omega' t}$, где $\omega' = \omega_k - \omega_l \neq 0$, интеграл которой по ее периоду равен нулю. Таким образом, весь J распадется на сумму функционалов от каждой гармоники по отдельности, для вычисления которых удобно воспользоваться следующей формулой: если даны произвольные $\xi, \eta \in \mathbf{C}^2$ и действительное $\omega = \frac{T}{2\pi}k$ ($k-$ целое), то для комплексных функций

$$z = \xi\, e^{i\omega t}, \qquad y = \eta\, e^{i\omega t},$$

справедливо равенство

$$\int_0^T (\operatorname{Re} z, \operatorname{Re} y)\, dt \;=\; \frac{1}{2}\operatorname{Re}\int_0^T \langle z, y\rangle\, dt = \;\frac{T}{2}\;\operatorname{Re}\;\langle \xi, \eta\rangle, \tag{35}$$

где круглыми скобками обозначено покомпонентное скалярное произведение двумерных векторов, а угловыми — комплексное скалярное произведение. (Это равенство устанавливается непосредственной проверкой.)

Так как коэффициенты при всех гармониках независимы друг от друга, то для каждой гармоники с номером $k \geq 1$ получим требование

$$J = \operatorname{Re}\int_0^T \Big(\omega_k^2\, |\xi_k|^2 - 2b\, i\omega_k\, (P\xi_k, \bar{\xi}_k) + (Q\xi_k, \bar{\xi}_k)\Big)\; dt \;\geq\; 0. \tag{36}$$

Изучим это неравенство (опуская индекс k). Пусть $\xi = f + ih$, где $f, h \in \mathbb{R}^2$. Тогда подинтегральное выражение в (36) запишется в виде

$$\omega^2\left(|f|^2 + |h|^2\right) - 2i\omega b\left(P(f+ih), (f-ih)\right) + \left(Q(f+ih), (f-ih)\right),$$

поэтому неравенство (36) для его действительной части означает:

$$\omega^2(f^2 + h^2) - 4\omega b\,(Pf, h) + (Qf, f) + (Qh, h)\;\geq 0. \tag{37}$$

Это неравенство должно выполняться для любых векторов $f, h \in \mathbb{R}^2$ и для любого ω, кратного $T/2\pi$.

Теперь заметим, что поскольку T произвольно, то неравенство (37) должно выполняться $\forall\, \omega > 0$. Если к тому же заменим в этом неравенстве ω на $-\omega$, а f на $-f$, то неравенство не изменится, поэтому оно должно выполняться вообще для всех $\omega \in \mathbb{R}$. Рассмотрение гармоники $k = 0$ приводит к очевидному требованию $Q \geq 0$, которое уже содержится в (37) (надо взять $\omega = 0$, $h = 0$).

Итак, мы приходим к следующему утверждению.

Предложение 2. Для неотрицательности J на всех финитных функциях необходимо и достаточно, чтобы для любых векторов $f, h \in \mathbb{R}^2$ и любого $\omega \in \mathbb{R}$ выполнялось неравенство (37).

Доказательство. a) Достаточность. Пусть выполнено (37). Тогда для любого $T > 0$ и любой гармоники на отрезке $[0, T]$ выполнено (36). Следовательно, для любой конечной суммы таких гармоник $J \geq 0$, а тогда в силу упоминавшейся плотности конечных сумм в $L_2[0, T]$ получаем $J \geq 0$ и для любой пары $(x, u) \in L_2[0, T]$ (в частности, для таких пар, у которых $x(T) = 0$). Из произвольности T отсюда следует, что $J \geq 0$ для любых финитных $(x, u) \in L_2\,[0, \infty)$.

b) Необходимость докажем от противного. Пусть (37) не выполнено при некоторых ω, f, h. Тогда для соответствующей гармоники (x, u) на ее периоде $[0, T]$ получим нарушение (36), т.е. $J(x, u) = -\alpha < 0$, а на ее кратном периоде $[1, NT + 1]$ при любом N будет $J(x, u) = -\alpha N$. Соединив теперь $x(1)$ линейно с $x(0) = 0$ на отрезке $[0, 1]$, а $x(NT + 1) = x(1)$ с $x(NT + 2) = 0$ на отрезке $[NT + 1,\, NT + 2]$, получим лишь конечный добавок в интеграл (из-за ограниченности $x(t)$), так что при большом N будем иметь финитную пару $(\hat{x}, \hat{u})$, на которой $J < 0$, ч.т.д.

Описанный прием изучения интегрального квадратичного функционала на $[0, \infty)$ путем перехода к разложению x, u в ряд Фурье использовался в работах [5, 6, 7]; в них также был получен и критерий (37).

Проанализируем этот критерий. Так как относительно ω мы имеем квадратный трехчлен, то, взяв его дискриминант, получаем, что (37) эквивалентно неравенству

$$4b^2(Pf, h)^2 \;\le\; ((Qf, f) + (Qh, h))\,(f^2 + h^2), \tag{38}$$

которое должно выполняться для любых $f, h \in \mathbb{R}^2$. В такой форме критерий неотрицательности J был получен в [3]. Его недостатком является то, что (38) есть неравенство четвертой степени относительно произвольной пары векторов $f, h,$ и не очень ясно, как его исследовать дальше.

Покажем, что на самом деле (38) эквивалентно (12). Как и прежде, без нарушения общности считаем $Q = \operatorname{diag}(q_1, q_2)$. В этом же базисе представим векторы f, h в координатном виде: $f = (f_1, f_2)$, $h = (h_1, h_2)$. Тогда (38) приобретает вид:

$$4b^2(f_1h_2 - f_2h_1)^2 \;\le\; \Big(q_1(f_1^2 + h_1^2) + q_2(f_2^2 + h_2^2)\Big)\,(f_1^2 + h_1^2 + f_2^2 + h_2^2). \tag{39}$$

Если ввести векторы $x = (f_1, h_1)$, $y = (f_2, y_2)$, то неравенство (39) перепишется так:

$$4b^2(Px, y)^2 \;\le\; (q_1x^2 + q_2y^2)\,(x^2 + y^2). \tag{40}$$

Ясно, что при любых фиксированных $|x|$, $|y|$ максимум левой части достигается при $x \perp y$ (т.е. реализуется самый худший случай); тогда $|(Px, y)| = |x|\cdot|y|$, поэтому (40) эквивалентно неравенству

$$4b^2x^2\,y^2 \;\le\; (q_1x^2 + q_2y^2)\,(x^2 + y^2) \tag{41}$$

относительно скалярных величин x^2 и y^2. При $|x| = 0$ или $|y| = 0$ оно выполнено тривиально, поэтому достаточно проверить его при $|x| > 0$, $|y| > 0$. Тогда, положив $|y|^2 = \alpha|x|^2$, получаем, что $\forall\,\alpha > 0$ должно выполняться неравенство

$$4b\alpha \le (q_1 + q_2\alpha)(1 + \alpha),$$

поделив которое на α, получаем

$$4b \;\le\; (q_1 + q_2) + \frac{q_1}{\alpha} + q_2\alpha. \tag{42}$$

Минимум правой части (опять – самый худший случай) достигается, как известно, при при $q_1/\alpha = q_2\alpha$, т.е. при $\alpha = \sqrt{q_1/q_2}$. Тогда (42) превращается в неравенство

$$4b^2 \;\le\; (q_1 + q_2) + 2\sqrt{q_1q_2}. \tag{43}$$

Но правая часть здесь есть полный квадрат: $(\sqrt{q_1}+\sqrt{q_2})^2$, поэтому (43) эквивалентно тому, что $2|b| \le \sqrt{q_1} + \sqrt{q_2}$, а это и есть полученное выше неравенство (12).

8. Обоснование метода. Остановимся на обосновании предложенного в п. 3 приема. Почему добавление под интеграл выражения $\frac{d}{dt}(Sx,x)$ приводит к успеху? Обратим внимание, что для задачи минимизации функционала (7) на решениях системы (6) функция $\varphi(x) = (Sx,x)$ является т.н. функцией Кротова. Напомним, что в общем случае функция Кротова – это такая функция $\varphi(x,t)$, для которой подинтегральное выражение при добавлении $\frac{d}{dt}\varphi(x,t)$ достигает минимума по всем x,u, не связанным между собой уравнением (6), на исследуемой траектории $x^0(t)$, $u^0(t)$. Если такая функция (гладкая) существует, то легко устанавливается, что данная траектория доставляет сильный минимум в задаче. (Подробнее см. [9, 10].) Таким образом, существование функции Кротова есть достаточное условие сильного минимума. При изучении функционала (7) мы фактически берем в качестве исследуемой траектории $x^0(t) \equiv u^0(t) \equiv 0$, и минимальность этой траектории означает, что $J \geq 0$ на подпространстве (6).

Так как функционал (7) квадратичный, то и функцию Кротова естественно искать в виде квадратичной формы: $\varphi(x) = (Sx,x)$. Нетривиальный (и в общем случае мало исследованный) вопрос состоит в том, необходимо ли существование функции Кротова для оптимальности данной траектории? В нашем случае: *необходимо ли существование требуемой квадратичной формы* (Sx,x) *для неотрицательности функционала (7)?* Другими словами, пусть не существует симметричной матрицы S, для которой подинтегральное выражение в (8) неотрицательно для независимых x,u (т.е. не существует S, для которой $M \geq 0$). Почему тогда найдутся x,u, удовлетворяющие (6), для которых $J < 0$? В п. 3 этот факт был установлен предъявлением конкретной пары (x,u) для двумерного случая. А.А.Милютин показал [11], что он справедлив и для случая произвольной линейной системы $\dot{x} = Ax + Bu$ с постоянными коэффициентами при произвольных размерностях x,u. Этот факт интересен тем, что он представляет собой довольно редкий случай, когда можно утверждать, что существование функции Кротова является необходимым для оптимальности. Мы приведем здесь его доказательство для случая простейшей управляемой системы $\dot{x} = u$, $x \in \mathbb{R}^n$ (с произвольным n), в котором оно технически гораздо проще, чем в общем.

Будем рассматривать квадратичный функционал

$$J = \int_0^\infty \Big(u^2 + (Vx,u) + (Qx,x) \Big)\, dt \tag{44}$$

на подпространстве $\mathcal{L}$:

$$\dot{x} = u, \qquad x(0) = 0, \tag{45}$$

где $x,u \in \mathbb{R}^n$, матрица $Q-$ симметричная, $V-$ кососимметричная (симметричная часть V дает в интеграле ноль). Нас интересует вопрос, когда $J \geq 0$ на подпространстве $\mathcal{L}$.

Возьмем, как и раньше, произвольную симметричную матрицу S и, добавив под интеграл $\frac{d}{dt}(Sx,x)$, получим

$$J = \int_0^\infty \Big([u + (S+V)x]^2 + (M(S)x,x) \Big)\, dt, \tag{46}$$

$$\text{где} \qquad (M(S)x,x) = (Qx,x) - (Sx+Vx)^2, \tag{47}$$

$M(S)-$ симметричная матрица.

Следуя А.А.Милютину, обозначим через $\lambda(S)$ минимальное собственное значение матрицы $M(S)$, т.е.

$$\lambda(S) \;=\; \min_{|x|=1}\,(M(S)x,x).$$

Положим $\lambda^0 = \sup \lambda(S)$ по всем симметричным S. Если $\lambda^0 > 0$, то $M(S) > 0$ для некоторой S, поэтому $J \ge 0$. Если $\lambda^0 = 0$, то $\forall\, \varepsilon > 0$ рассмотрим функционал $J_\varepsilon \;=\; J + \int \varepsilon(x,x)\,dt$, для которого $Q_\varepsilon \;=\; Q + \varepsilon E$, и поэтому $\lambda^0_\varepsilon \;=\; \lambda^0 + \varepsilon \;>\; 0$. Следовательно, $J_\varepsilon \ge 0$, а тогда, переходя к пределу, получим и $J \ge 0$ на $\mathcal{L}$.

Осталось рассмотреть случай $\lambda^0 < 0$. Наша цель — показать, что в этом случае найдется $x \in \mathcal{L}$, для которого $J(x) < 0$. Тем самым будет доказана

Теорема 3 (А.А.Милютин). *$J \ge 0$ на $\mathcal{L}$ тогда и только тогда, когда $\lambda^0 \ge 0$.*

Эта теорема служит обоснованием предложенной выше процедуры поиска подходящей матрицы S. Доказательство использует следующее свойство матриц, установленное Милютиным. Пусть дана произвольная $n \times n-$ матрица R.

Лемма 6 [11]. Следующие два условия эквивалентны:
(a) для любой симметричной матрицы S

$$\max_{|x|=1}\,(Rx,Sx) \;\ge\; 0\,; \tag{48}$$

(b) матрица R имеет собственное значение с нулевой вещественной частью.

Доказательство. Покажем, что $(b) \Longrightarrow (a)$. Если $\lambda \;=\; 0$ есть собственное число матрицы R, то существует такой вектор x_0, что $|x_0| = 1$ и $Rx_0 = 0$. Тогда $\forall\, S$ имеем $(Rx_0, Sx_0) = 0$, поэтому (48) выполнено.

Пусть $\lambda \;=\; i\omega \;\ne\; 0$ есть собственное число матрицы R. Тогда, как известно, уравнение $\dot{x} \;=\; Rx$ имеет периодическое решение вида $x^0(t) \;=\; f\sin\omega t + h\cos\omega t$, где векторы $f, h \in \mathbb{R}^n$ линейно независимы, и следовательно, всюду $x^0(t) \;\ne\; 0$. Возьмем произвольную симметричную матрицу S и рассмотрим функцию $m(t) = (Sx^0(t), x^0(t))$. В силу ее периодичности она имеет точку максимума t_* . В этой точке $\frac{d}{dt}m(t_*) = 2(Sx^0(t_*), Rx^0(t_*)) = 0$, и так как $x^0(t_*) \ne 0$, то отсюда следует (48), ч.т.д.

Импликацию $(a) \Longrightarrow (b)$ докажем от противного. Пусть все собственные значения λ матрицы R имеют $\operatorname{Re}\lambda \ne 0$. Нам надо показать, что найдется симметричная S, для которой (48) нарушается. Это эквивалентно тому, что должна существовать квадратичная форма $\varphi \;=\; (Sx,x)$, которая на любом ненулевом решении системы $\dot{x} \;=\; Rx$ имеет отрицательную производную: $\frac{d}{dt}\varphi(x(t)) \;=\; (Sx,Rx) \;<\; 0$, т.е. является функцией Ляпунова этой системы. Но существование квадратичной функции Ляпунова не зависит от базиса, в котором рассматривается система, поэтому будем искать эту функцию в том базисе, где матрица R приводится к вещественной жордановой форме. Ясно, что достаточно рассмотреть случай одной вещественной жордановой клетки (т.к. каждая клетка соответствует инвариантному подпространству матрицы R), т.е. считаем, что R есть клетка, соответствующая паре комплексно-сопряженных собственных значений $\lambda,\ \bar{\lambda}$. Кроме того, можно считать, что $\mathrm{Re}\,\lambda < 0$ (иначе заменим R на $-R$ и S на $-S$). Как известно из теории обыкновенных дифференциальных уравнений, в этом случае система $\dot{x} = Rx$ имеет квадратичную функцию Ляпунова $f(x) = (Sx, x)$. Теперь осталось сложить квадратичные формы, соответствующие всем жордановым клеткам, и вернуться обратно в исходный базис. Лемма доказана.

Нам потребуется также следующий нетривиальный факт. Обозначим через Σ единичную сферу в $\mathbb{R}^n$.

Лемма 7 [11]. Пусть ненулевое подпространство $\Gamma_0 \subset \mathbb{R}^n$ и $n \times n-$ матрица R таковы, что для любой симметричной $n \times n-$ матрицы S

$$\max_{x \in \Gamma_0 \cap \Sigma} (Rx, Sx) \geq 0. \tag{49}$$

Тогда существует ненулевое подпространство $\Gamma \subset \Gamma_0$, инвариантное относительно R (т.е. $R\Gamma \subset \Gamma$), для которого также выполнено (49).

Доказательство. Пусть Γ_1 есть ортогональное дополнение к Γ_0, т.е. $\mathbb{R}^n = \Gamma_0 \oplus \Gamma_1$. Обозначим через π_0 и π_1 ортогональные проекции на Γ_0 и Γ_1 соответственно. Для любого оператора $A : \mathbb{R}^n \to \mathbb{R}^n$ будем рассматривать операторы $A_0 = \pi_0 A$ и $A_1 = \pi_1 A$, так что $A = A_0 + A_1$. При этом всегда выполняется разложение

$$(Rx, Sx) = (R_0 x, S_0 x) + (R_1 x, S_1 x). \tag{50}$$

Заметим, что оператор A совпадает на подпространстве Γ_0 с некоторым симметричным оператором S тогда и только тогда, когда A_0- симметричный оператор на Γ_0; при этом A_1 может быть совершенно произвольным. (Удобно представлять себе Γ_0 и Γ_1 как координатные подпространства.)

Определим подпространство $L = \{x \in \Gamma_0 \mid Rx \in \Gamma_0)\}$. Если $L = \Gamma_0$, т.е. Γ_0 инвариантно для R, то все доказано, поэтому надо рассмотреть случай $L \neq \Gamma_0$.

Установим, что всегда $L \neq \{0\}$. Другими словами, что оператор R_1 имеет нетривиальное ядро на Γ_0. Если это не так, т.е. $R_1 x \neq 0$ на $\Gamma_0 \cap \Sigma$, то возьмем симметричную S, которая на Γ_0 имеет $S_0 x = 0$, $S_1 x = -R_1 x$. Для нее на $\Gamma_0 \cap \Sigma$ получим

$$(Rx, Sx) = -(R_1 x, S_1 x) = -|R_1 x|^2 < 0,$$

что противоречит условию леммы.

Итак, $L-$ ненулевое подпространство в Γ_0, и мы считаем, что $L \neq \Gamma_0$. Мы утверждаем, что для L также выполнено (49), т.е. для любой симметричной матрицы S

$$\max_{x \in L \cap \Sigma} (Rx, Sx) \geq 0. \tag{51}$$

Допустим, это не так. Тогда существует симметричная матрица S и число $\alpha_0 > 0$ такие, что на $L \cap \Sigma$

$$(Rx, Sx) \leq -\alpha_0 < 0. \tag{52}$$

Разложим Γ_0 в ортогональную сумму: $\Gamma_0 = L \oplus H$, и для $x \in \Gamma_0$ будем писать $x = x_L + x_H$. При этом по определению $RL \subset \Gamma_0$ и $RH \cap \Gamma_0 = \{0\}$.

Возьмем теперь симметричную матрицу $\hat{S}$ таким образом, что на Γ_0

$$\hat{S}_0 = S_0, \qquad \hat{S}_1 = -NR_1 \ .$$

Для нее, согласно (50), на Γ_0 будет выполнено:

$$(Rx, \hat{S}x) = (R_0x, S_0x) - N|R_1x|^2. \tag{53}$$

Так как $RL \subset \Gamma_0$, то $R_1L = 0$, поэтому для $x = x_L + x_H$ получаем $R_1x = R_1x_H$, а так как $RH \cap \Gamma_0 = \{0\}$, то $|R_1x_H|^2 \geq \alpha_1 \, |x_H|^2$ при некотором $\alpha_1 > 0$.

Распишем для $x = x_L + x_H \in \Gamma_0$ первый член в правой части (53):

$$(R_0x, S_0x) = \ (R_0x_L, S_0x_L) + (R_0x_L, S_0x_H) + (R_0x_H, S_0x_L) + (R_0x_H, S_0x_H).$$

Согласно (52), здесь $(R_0x_L, S_0x_L) \leq -\alpha_0 \, |x_L|^2$, поэтому для левой части (53) имеем на Γ_0 оценку

$$(Rx, \hat{S}x) \leq -\alpha_0 \, |x_L|^2 + 2\beta \, |x_L| \cdot |x_H| + \gamma \, |x_H|^2 - N\alpha_1 \, |x_H|^2$$

с некоторыми константами β, γ. Справа здесь стоит двумерная квадратичная форма относительно $|x_L|$, $|x_H|$. Ясно, что при достаточно большом N она отрицательно определена, откуда следует нарушение (51) для $\hat{S}$.

Итак, мы показали, что если подпространство Γ_0 не инвариантно относительно R, то в нем существует ненулевое подпространство L меньшей размерности, для которого по-прежнему выполнено (49), т.е. (51). Продолжая этот процесс, за конечное число шагов придем к инвариантному подпространству L. Лемма доказана.

Обратимся теперь к определению λ^0.

Лемма 8. $\sup \lambda(S)$ по всем симметричным S достигается, т.е. $\exists\, S$, для которой $\lambda^0 = \lambda(S)$.

Доказательство. Априори ясно, что $\lambda^0 > -\infty$. Пусть последовательность S_k такова, что $\lambda(S_k) \to \lambda^0$. Если последовательность норм $||S_k||$ ограничена, то можно считать $S_k \to S_0$, и тогда, очевидно, справедливо предельное равенство $\lambda(S_0) = \lambda^0$ (так как минимум квадратичной формы на единичной сфере непрерывно зависит от ее коэффициентов). Предположим теперь, что $||S_k|| \to \infty$ (на некоторой подпоследовательности). При этом найдутся векторы x_k такие, что $|x_k| = 1$ и $|S_kx_k| \to \infty$. Тогда согласно (47) главный член в $(M(S_k)x_k, x_k)$ есть $-|S_kx_K|^2 \to -\infty$, и следовательно,

$$\lambda(S_k) = \ \min_{|x|=1} (M(S_k)x, x) \leq \ (M(S_k)x_k, x_k) \to -\infty,$$

что противоречит стремлению $\lambda(S_k) \to \lambda^0$. Лемма доказана.

Теперь мы можем дать

Доказательство теоремы 3. Как уже говорилось, достаточно рассмотреть случай $\lambda^0 < 0$.

Для любой симметричной матрицы S существует ненулевое подпространство $\Gamma(S) \subset \mathbb{R}^n$ (являющееся инвариантным для матрицы S) такое, что

$$Arg \min_{x\in\Sigma} (M(S)x, x) = \Gamma(S) \cap \Sigma.$$

(Действительно, квадратичная форма $(M(S)x, x) - \lambda(S)(x, x) \geq 0$ на $\mathbb{R}^n$ и имеет нетривиальное подпространство нулей; это и есть требуемое $\Gamma(S)$.)

Возьмем теперь любую S_0, на которой достигается $\max \lambda(S)$. Она, таким образом, является решением задачи

$$\lambda(S) = \min_{x\in\Sigma} (M(S)x, x) \ \to \max_{S},$$

в которой x играет роль параметра.

Обозначим $\Gamma_0 = \Gamma(S_0)$, и введем матрицу $R = S_0 + V$, так что

$$M(S_0) = \ (Qx, x) - (Rx)^2. \tag{54}$$

Основной пункт в доказательстве теоремы 3 составляет следующий факт.

Утверждение. Для любой симметричной матрицы $\bar{S}$

$$\max_{x\in\Gamma_0\cap\Sigma} (Rx, \bar{S}x) \geq 0. \tag{55}$$

Доказательство вытекает из формулы производной по направлению функции минимума. Рассмотрим симметричную матрицу $S_\varepsilon = S_0 + \varepsilon\bar{S}$ при малом $\varepsilon > 0$. Ей соответствует значение

$$\lambda(S_\varepsilon) = \min_{x\in\Sigma} (M(S_\varepsilon)x, x),$$

которое есть функция от ε. Так как S_0 есть точка максимума $\lambda(S)$, то всегда $\lambda(S_\varepsilon) \leq \lambda(S_0)$, поэтому правая производная (при $\varepsilon \to 0+$) $\frac{d}{d\varepsilon}\lambda(S_\varepsilon) \leq 0$. Вычислим эту производную.

Напомним, что если $\varphi(\varepsilon) = \min_{x\in K} \Phi(\varepsilon, x)$, где $K-$ компакт, а $\Phi-$ гладкая функция, то

$$\varphi'(0+) = \min_{x\in K_0} \Phi'_\varepsilon(0, x),$$

где $K_0 = Arg \min\, \Phi(0, x) \mid x \in K$.

В нашем случае $\Phi(\varepsilon, x) = (M(S_\varepsilon)x, x)$, $\quad Arg \min \{\Phi(0, x) \mid x \in \Sigma\} = \Gamma_0 \cap \Sigma$, и согласно (54) $\Phi'_\varepsilon(0, x) = -2(Rx, \bar{S}x)$. Следовательно,

$$\frac{d}{d\varepsilon}\lambda(S_\varepsilon)\bigg|_{\varepsilon=0+} = \ \varphi'(0+) = \min_{x\in\Gamma_0\cap\Sigma} (-2Rx, \bar{S}x) \leq 0,$$

что эквивалентно (55).

Итак, (55) установлено. По лемме 7 существует ненулевое подпространство $\Gamma \subset \Gamma_0$ такое, что для любой симметричной $\bar{S}$

$$\max_{x\in\Gamma\cap\Sigma} (Rx, \bar{S}x) \geq 0, \tag{56}$$

и при этом $R\Gamma \subset \Gamma$. Отсюда вытекает, что можно перейти к сужению R на подпространство Γ, т.е. к оператору $R: \Gamma \to \Gamma$, а неравенство (56) выполняется для любой матрицы $\bar{S}$, симметричной на этом подпространстве. Тогда по лемме 6 оператор R имеет на Γ собственное значение $\lambda = i\omega$ с нулевой вещественной частью.

Последнее означает, что в подпространстве Γ существует ненулевое решение уравнения $\dot{x} = -Rx$, а на этом решении первый квадрат в (46) по определению равен нулю. (Напомним, что у нас $R = S_0 + V$.) Если $\omega = 0$, это решение есть $x(t) \equiv x_0 \in \Gamma$; если $\omega \neq 0$, это решение есть $x(t) = f \sin\omega t + h\cos\omega t$, где $f, h \in \Gamma$. Так как на подпространстве Γ_0 (и тем более на Γ) $(Mx, x) = \lambda^0 < 0$, то на любом из этих решений можно набрать сколь угодно большой отрицательный интеграл от второго члена в (46). Повторяя теперь рассуждения из доказательства теоремы 1, получим финитную $\hat{x}(t)$, на которой $J(\hat{x}) < 0$. Теорема 3 доказана.

Заключение. В данной работе мы рассмотрели простейший нетривиальный случай квадратичного функционала с вырожденным условием Лежандра, преобразовав его в функционал с "хорошими" коэффициентами, но на интервале $[0, \infty)$. Интегральные функционалы на полуоси (точнее, на пространстве с бесконечной мерой) по своим свойствам качественно отличаются от интегральных функционалов на отрезках (т.е. на пространствах с конечной мерой); в них по-прежнему присутствует особенность. Это связано с тем обстоятельством, что интегральные операторы на пространстве с бесконечной мерой не являются, вообще говоря, вполне непрерывными.

Точные формулы для неотрицательности рассматриваемого функционала нам удалось получить лишь в двумерном случае. Уже для трехмерного случая этот вопрос пока открыт.

Настоящая работа выполнялась в рамках семинара под руководством А.А.Милютина. Автор благодарен его участникам за полезные обсуждения.

Список литературы

[1] Барабанов Н.Е., Гелиг А.Х., Леонов Г.А., Лихтарников А.Л., Матвеев А.С., Смирнова В.Б., Фрадков А.Л. Частотная теорема (лемма Якубовича-Калмана) в теории управления, – Автоматика и телемеханика, 1996, №10.

[2] Матвеев А.С., Якубович В.А. Невыпуклые задачи глобальной оптимизации в теории управления. – Итоги науки и техники, Соврем. математика и ее приложения. Тематические обзоры. ВИНИТИ, 1998, т. 60, с. 128–175.

[3] Арутюнов А.В. Условия экстремума. Анормальные и вырожденные задачи. // М.: Факториал, 1997, 254 с.

[4] Харди Г.Г., Литтльвуд Д.Е., Полиа Г. Неравенства.// М., ИЛ, 1948.

[5] Coppel W.A. Linear-quadratic optimal control. – Proc. of Royal Soc. Edinburgh, Sec. A, 1974/75, v. 73, p. 271–289.

[6] Bittanti S., Fronza G., Guardabassi G., Periodic control: A frequency domain approach. – IEEE Trans. Automat. Control, 1973, v. AC-18, p. 33–38.

[7] Guardabassi G., Locatelly A., Rinaldi S., Status of periodic optimization of dynamical systems. – JOTA, 1974, v. 12, no. 1, p. 1–20.

[8] Hestenes, M.R. Singular quadratic variational problems. – JOTA, 1983, v. 41, no. 1, p. 123–137.

[9] Кротов В.Ф., Гурман В.И. Методы и задачи оптимального управления. // М., Наука, 1973.

[10] Дмитрук А.В. О необходимости достаточных условий оптимальности кротовского типа. – Автоматика и телемеханика, 1997, №10, с. 3–17.

[11] Милютин А.А. Об условиях неотрицательности интегральных квадратичных форм с постоянными коэффициентами, определенных на полуоси. – Мат. сборник, 2002, т. 193, №4.

Опубликована в сб. *Итоги науки и техники. Современная математика и ее приложения, М., 2002, т. 110, 49–75.*

Центральный экономико-математический институт РАН,
117418, Москва, Нахимовский пр-т, 47.
E-mail: dmitruk@member.ams.org