ПОИСК НАИМЕНЬШЕГО РАССТОЯНИЯ МЕЖДУ ДВУМЯ ЭЛЛИПСОИДАМИ*

А. А. Чумаков
andrew1991.spb@gmail.com

Г. Ш. Тамасян
g.tamasyan@spbu.ru

27 марта 2014 г.

Аннотация. В работе рассматривается задача нахождения ближайших точек между двумя эллипсоидами [1–5]. Данная проблема условной оптимизации сводится к безусловной с помощью теории точных штрафных функций [6–8]. Построенная точная штрафная функция принадлежит к классу гиподифференцируемых функций [9]. Для решения поставленной задачи применяется хорошо известный и эффективный метод гиподифференциального спуска [5, 9]. Также разработан новый двухшаговый гипоградиентный метод. Приведены результаты численных экспериментов, демонстрирующие более высокую скорость сходимости нового метода.

1. Формулировка задачи

В пространстве $\mathbb{R}^n$ требуется найти ближайшие точки между двумя эллипсоидами Ω_1 и Ω_2, где

$$\Omega_k = \left\{ \xi \in \mathbb{R}^n \;\middle|\; h_k(\xi) := \frac{1}{2}\xi^T A_k \xi + \xi^T b_k + \sigma_k = 0 \right\}, \quad k = 1, 2, \tag{1}$$

A_k — постоянные $(n \times n)$-матрицы, симметричные и положительно определенные, b_k — постоянные n-мерные вектор-столбцы, σ_k — вещественные константы.

Пусть x_1, $x_2 \in \mathbb{R}^n$. Поставленную проблему будем рассматривать как задачу условной оптимизации $\|x_1 - x_2\| := \left[\sum_{j=1}^{n} (x_1^j - x_2^j)^2 \right]^{\frac{1}{2}} \longrightarrow \min\limits_{\substack{x_1 \in \Omega_1 \\ x_2 \in \Omega_2}}$.

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: `http://www.apmath.spbu.ru/cnsa/`

2. Штрафная функция

Пусть $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ векторы из $\mathbb{R}^{2n}$. На множестве $\mathbb{R}^{2n}$ введем в рассмотрение евклидову норму

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\|x_1\|^2 + \|x_2\|^2}$$

и метрику $\rho(x, y) = \|x - y\|$.

Введем функцию $f(x) = \dfrac{1}{2}\|x_1 - x_2\|^2$ и множество

$$X = \left\{ x \in \mathbb{R}^{2n} \;\middle|\; h_1(x_1) = 0,\ h_2(x_2) = 0 \right\}.$$

Множество X можно представить в эквивалентном виде

$$X = \left\{ x \in \mathbb{R}^{2n} \;\middle|\; \varphi(x) = 0 \right\},$$

где

$$\varphi(x) = |h_1(x_1)| + |h_2(x_2)|.$$

Пусть $\lambda \geqslant 0$ фиксировано. Введем функцию

$$\Phi_\lambda(x) = f(x) + \lambda\varphi(x).$$

Функция $\Phi_\lambda(x)$ называется *штрафной функцией*, а число λ — *штрафным параметром.*

Покажем, что для достаточно большого λ функция Φ_λ является точной штрафной функцией, т. е. что для достаточно большого λ любая точка глобального минимума штрафной функции Φ_λ является решением исходной задачи.

2.1 Классическая вариация функции φ и ее свойства

Изучим подробней свойства функции

$$\varphi(x) = |h_1(x_1)| + |h_2(x_2)|,$$

где $h_k(x_k) = \dfrac{1}{2}x_k^T A_k x_k + x_k^T b_k + \sigma_k$, $k = 1, 2$.

Пусть $x \in \mathbb{R}^{2n}$ фиксировано, $\varepsilon > 0$. Выберем произвольное $v \in \mathbb{R}^{2n}$. Положим

$$x_\varepsilon = x + \varepsilon v = \begin{pmatrix} x_1 + \varepsilon v_1 \\ x_2 + \varepsilon v_2 \end{pmatrix}. \tag{2}$$

Справедливо равенство

$$h_k(x_k+\varepsilon v_k)=h_k(x_k)+\varepsilon v_k^T\left(A_kx_k+b_k\right)+\varepsilon^2\frac{1}{2}v_k^TA_kv_k,\quad k=1,2. \tag{3}$$

Если $x\in X$, то $h_1(x_1)=h_2(x_2)=0$, и из (3), получаем явный вид производной функции $\varphi(x)$ в точке x по направлению v:

$$\varphi'(x,v):=\lim_{\varepsilon\downarrow 0}\frac{\varphi(x+\varepsilon v)-\varphi(x)}{\varepsilon}=\sum_{k=1}^{2}\left|v_k^T\left(A_kx_k+b_k\right)\right|.$$

Несложно проверить, что это выражение можно переписать в виде

$$\varphi'(x,v)=\max_{\substack{w_k\in[-1,1]\\ k=1,2}}\left[w_1v_1^T\left(A_1x_1+b_1\right)+w_2v_2^T\left(A_2x_2+b_2\right)\right]. \tag{4}$$

Из (4) заключаем, что функция $\varphi(x)$ субдифференцируема в точке x, причем ее субдифференциал $\partial\varphi(x)$ равен

$$\partial\varphi(x)=\operatorname{co}\left\{\begin{pmatrix}w_1\left(A_1x_1+b_1\right)\\ w_2\left(A_2x_2+b_2\right)\end{pmatrix}\;\middle|\; w_k\in[-1,1],\ k=1,2\right\}.$$

Значит,

$$\varphi'(x,v)=\max_{W\in\partial\varphi(x)}\langle v,W\rangle.$$

Пусть теперь $x\notin X$, т. е. $\varphi(x)>0$. Рассмотрим производную функции $\varphi(x)$ в точке x по направлению v. Возможны следующие три случая.

1. Если $h_1(x_1)\neq 0$, $h_2(x_2)\neq 0$, то

$$\varphi'(x,v)=w_1v_1^T\left(A_1x_1+b_1\right)+w_2v_2^T\left(A_2x_2+b_2\right), \tag{5}$$

где $w_1=\operatorname{sign}h_1(x_1)$, $w_2=\operatorname{sign}h_2(x_2)$. Из (5) заключаем, что функция φ дифференцируема в точке x, причем ее градиент $\frac{\partial\varphi(x)}{\partial x}$ равен

$$\frac{\partial\varphi(x)}{\partial x}=\begin{pmatrix}\left(A_1x_1+b_1\right)\operatorname{sign}h_1(x_1)\\ \left(A_2x_2+b_2\right)\operatorname{sign}h_2(x_2)\end{pmatrix}.$$

2. Если $h_1(x_1)=0$, $h_2(x_2)\neq 0$, то

$$\varphi'(x,v)=\max_{w_1\in[-1,1]}\left[w_1v_1^T\left(A_1x_1+b_1\right)+w_2v_2^T\left(A_2x_2+b_2\right)\right], \tag{6}$$

где $w_2=\operatorname{sign}h_2(x_2)$. Из (6) заключаем, что функция φ субдифференцируема в точке x, причем ее субдифференциал $\partial\varphi(x)$ равен

$$\partial\varphi(x)=\operatorname{co}\left\{\begin{pmatrix}-\left(A_1x_1+b_1\right)\\ \left(A_2x_2+b_2\right)\operatorname{sign}h_2(x_2)\end{pmatrix},\begin{pmatrix}A_1x_1+b_1\\ \left(A_2x_2+b_2\right)\operatorname{sign}h_2(x_2)\end{pmatrix}\right\}.$$

3. Если $h_1(x_1) \neq 0$, $h_2(x_2) = 0$, то

$$\varphi'(x, v) = \max_{w_2 \in [-1,1]} \left[w_1 v_1^T (A_1 x_1 + b_1) + w_2 v_2^T (A_2 x_2 + b_2) \right], \tag{7}$$

где $w_1 = \operatorname{sign} h_1(x_1)$. Из (7) заключаем, что функция φ субдифференцируема в точке x, причем ее субдифференциал $\partial\varphi(x)$ равен

$$\partial\varphi(x) = \operatorname{co}\left\{ \begin{pmatrix} (A_1 x_1 + b_1) \operatorname{sign} h_1(x_1) \\ -(A_2 x_2 + b_2) \end{pmatrix}, \begin{pmatrix} (A_1 x_1 + b_1) \operatorname{sign} h_1(x_1) \\ A_2 x_2 + b_2 \end{pmatrix} \right\}.$$

2.2 Точная штрафная функция

Напомним, что скоростью наискорейшего спуска функции φ в точке x из $\mathbb{R}^{2n}$ называется величина

$$\varphi^{\downarrow}(x) = \liminf_{\substack{y \in \mathbb{R}^{2n} \\ y \to x}} \frac{\varphi(y) - \varphi(x)}{\rho(y, x)}. \tag{8}$$

Справедлива следующая теорема.

ТЕОРЕМА 1. *Существуют $\delta > 0$ и $a > 0$ такие, что для любого $x \in X_\delta \setminus X$ справедливо неравенство*

$$\varphi^{\downarrow}(x) \leqslant -a < 0, \tag{9}$$

где $X_\delta = \left\{ x \in \mathbb{R}^{2n} \;\middle|\; \varphi(x) < \delta \right\}$.

Доказательство. Покажем справедливость соотношения (9) в каждом из трех вышерассматриваемых случаях (см. п. 2.1).

В случае 1 для любого $\alpha > 0$ положим $x_\alpha = x + \alpha v^*$, где

$$v^* = -\begin{pmatrix} (A_1 x_1 + b_1) \operatorname{sign} h_1(x_1) \\ (A_2 x_2 + b_2) \operatorname{sign} h_2(x_2) \end{pmatrix}.$$

Тогда, используя (3) и (5), имеем

$$\varphi(x_\alpha) = \varphi(x) + \alpha v^{*T} \begin{pmatrix} (A_1 x_1 + b_1) \operatorname{sign} h_1(x_1) \\ (A_2 x_2 + b_2) \operatorname{sign} h_2(x_2) \end{pmatrix} + o_1(\alpha) =$$
$$= \varphi(x) - \alpha H_1^*(x) + o_1(\alpha), \tag{10}$$

где $\dfrac{o_1(\alpha)}{\alpha} \underset{\alpha \downarrow 0}{\longrightarrow} 0$ и $H_1^*(x) = \left\| \begin{pmatrix} A_1 x_1 + b_1 \\ A_2 x_2 + b_2 \end{pmatrix} \right\|^2$. Заметим, что найдется такое $\delta_1 > 0$, при котором $H_1^*(x) > 0$ для всех $x \in X_{\delta_1} \setminus X$.

Вычислим
$$\rho(x_\alpha, x) = \alpha \left\| \begin{pmatrix} A_1x_1 + b_1 \\ A_2x_2 + b_2 \end{pmatrix} \right\|. \tag{11}$$
Из (8) имеем
$$\varphi^{\downarrow}(x) \leqslant \liminf_{\alpha\downarrow 0} \frac{\varphi(x_\alpha) - \varphi(x)}{\rho(x_\alpha, x)}. \tag{12}$$
Подставляя (10), (11) в (12), получим
$$\varphi^{\downarrow}(x) \leqslant - \left\| \begin{pmatrix} A_1x_1 + b_1 \\ A_2x_2 + b_2 \end{pmatrix} \right\| < 0. \tag{13}$$

В случае 2 для любого $\alpha > 0$ положим $x_\alpha = x + \alpha v^*$, где
$$v^* = -\begin{pmatrix} \mathbb{O}_n \\ (A_2x_2 + b_2)\operatorname{sign} h_2(x_2) \end{pmatrix}.$$
Тогда, используя (3) и (6), имеем
$$\varphi(x_\alpha) = \varphi(x) - \alpha\|A_2x_2 + b_2\|^2 + o_2(\alpha) = \varphi(x) - \alpha H_2^*(x) + o_2(\alpha), \tag{14}$$
где $\dfrac{o_2(\alpha)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0$. Заметим, что найдется такое $\delta_2 > 0$, при котором $H_2^*(x) > 0$ для всех $x \in X_{\delta_2} \setminus X$.

Далее,
$$\rho(x_\alpha, x) = \alpha \left\| A_2x_2 + b_2 \right\|. \tag{15}$$
Подставляя (14), (15) в (12), получим
$$\varphi^{\downarrow}(x) \leqslant - \left\| A_2x_2 + b_2 \right\| < 0. \tag{16}$$

Наконец, рассмотрим случай 3. Как и в случае 2, для любого $\alpha > 0$ положим $x_\alpha = x + \alpha v^*$, где
$$v^* = -\begin{pmatrix} (A_1x_1 + b_1)\operatorname{sign} h_1(x_1) \\ \mathbb{O}_n \end{pmatrix}.$$
Тогда, используя (3) и (7), имеем
$$\varphi(x_\alpha) = \varphi(x) - \alpha\|A_1x_1 + b_1\|^2 + o_3(\alpha) = \varphi(x) - \alpha H_3^*(x) + o_3(\alpha), \tag{17}$$
где $\dfrac{o_3(\alpha)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0$. Заметим, что найдется такое $\delta_3 > 0$, при котором $H_3^*(x) > 0$ для всех $x \in X_{\delta_3} \setminus X$.

Далее,

$$\rho(x_\alpha, x) = \alpha \left\| A_1 x_1 + b_1 \right\|. \tag{18}$$

Подставляя (17), (18) в (12), получим

$$\varphi^{\downarrow}(x) \leqslant - \left\| A_1 x_1 + b_1 \right\| < 0. \tag{19}$$

Из полученных выше неравенств (13), (16) и (19) следует справедливость неравенства (9), где $\delta = \min\{\delta_1, \delta_2, \delta_3\}$, $a = \min\{a_1, a_2, a_3\}$,

$$a_1 = \inf_{x \in X_{\delta_1} \setminus X} \left\| \begin{pmatrix} A_1 x_1 + b_1 \\ A_2 x_2 + b_2 \end{pmatrix} \right\|, \quad a_2 = \inf_{x \in X_{\delta_2} \setminus X} \left\| A_2 x_2 + b_2 \right\|,$$

$$a_3 = \inf_{x \in X_{\delta_3} \setminus X} \left\| A_1 x_1 + b_1 \right\|.$$

□

Теперь нетрудно получить утверждение дающее, достаточное условие того, что функция $\Phi_\lambda(x) = f(x) + \lambda \varphi(x)$ является точной штрафной функцией.

ТЕОРЕМА 2. *Предположим, что существует $\lambda_0 < \infty$ такое, что для любого $\lambda \geqslant \lambda_0$ множество $\arg\min\limits_{x \in \mathbb{R}^{2n}} \Phi_\lambda(x)$ не пусто. Тогда существует такое $\lambda^* \geqslant \lambda_0$, что для любого $\lambda > \lambda^*$ множество $\arg\min\limits_{x \in \mathbb{R}^{2n}} \Phi_\lambda(x)$ не пусто и для всех $x_\lambda \in \arg\min\limits_{x \in \mathbb{R}^{2n}} \Phi_\lambda(x)$ будет*

$$x_\lambda \in X, \quad f(x_\lambda) = \inf_{x \in X} f(x),$$

т. е. $\Phi_\lambda(x)$ является точной штрафной функцией.

Доказательство следует из **теоремы 1** и **теоремы 3.4.2** из [6].

Справедлива следующая теорема о точках локального минимума штрафной функции Φ_λ.

ТЕОРЕМА 3 ([6]). *Если $x_* \in X$ — точка локального минимума функции f на множестве X в метрике ρ, то найдется $\lambda^* < \infty$, такое, что при $\lambda > \lambda^*$ точка x_* является точкой локального минимума функционала $\Phi_\lambda(x)$ на всем пространстве $\mathbb{R}^{2n}$ в той же метрике ρ.*

3. Классическая вариация функции $f(x)$

Для классической вариации (2) имеем

$$f(x_\varepsilon) = f(x) + \varepsilon v^T \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \frac{1}{2}\varepsilon^2 \left\| v_1 - v_2 \right\|^2. \tag{20}$$

Отсюда следует, что функция f дифференцируема в точке x, а функция

$$Q(x) = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} \tag{21}$$

является градиентом функции f в точке x.

4. Необходимые условия минимума

Пусть $x_* \in X$ — точка локального минимума функционала f на множестве X. Ранее было установлено, что найдется такое $\lambda^* < \infty$, что при $\lambda > \lambda^*$ точка x_* будет является точкой локального минимума функционала $\Phi_\lambda(x) = f(x) + \lambda\varphi(x)$ на всем пространстве $\mathbb{R}^{2n}$.

Зафиксируем произвольное $\lambda > \lambda^*$. Пусть $\varepsilon > 0$, выберем $v \in \mathbb{R}^{2n}$ и положим

$$x_\varepsilon = x_* + \varepsilon v. \tag{22}$$

Поскольку $\varphi(x_*) = 0$, то для вариации (22) из (4) и (20) получим

$$\begin{aligned} \Phi_\lambda(x_\varepsilon) = \Phi_\lambda(x_*) + \varepsilon\Big[& v^T Q(x_*) + \\ & + \lambda\Big\{ \left| v_1^T (A_1 x_1 + b_1) \right| + \left| v_2^T (A_2 x_2 + b_2) \right| \Big\}\Big] + o(\varepsilon), \end{aligned} \tag{23}$$

где $Q(x_*)$ — выражение (21) при $x = x_*$ и $\dfrac{o(\varepsilon)}{\varepsilon} \underset{\varepsilon\downarrow 0}{\longrightarrow} 0$. Соотношение (23) можно переписать в виде

$$\Phi_\lambda(x_\varepsilon) = \Phi_\lambda(x_*) + \varepsilon \max_{\substack{w_k \in [-1,1] \\ k=1,2}} v^T \left[Q(x_*) + \lambda \begin{pmatrix} w_1 (A_1 x_1 + b_1) \\ w_2 (A_2 x_2 + b_2) \end{pmatrix} \right] + o(\varepsilon). \tag{24}$$

ТЕОРЕМА 4 ([6]). *Для того, чтобы $x_* \in X$ была точкой глобального или локального минимума функции f на множестве X в метрике ρ, необходимо, чтобы выполнялось неравенство*

$$\Phi_\lambda^{\downarrow}(x_*) := \liminf_{\substack{y \in \mathbb{R}^{2n} \\ y \to x_*}} \frac{\Phi_\lambda(y) - \Phi_\lambda(x_*)}{\rho(y, x_*)} \geqslant 0.$$

Учитывая произвольность $v \in \mathbb{R}^{2n}$, из (24) и **теоремы 4** имеем необходимое условие минимума

$$\max_{\substack{w_k \in [-1,1] \\ k=1,2}} v^T \left[Q(x_*) + \lambda \begin{pmatrix} w_1 \left(A_1 x_1 + b_1\right) \\ w_2 \left(A_2 x_2 + b_2\right) \end{pmatrix} \right] \geqslant 0 \quad \forall\, v \in \mathbb{R}^{2n}. \tag{25}$$

ТЕОРЕМА 5. *Соотношение* (25) *эквивалентно следующему условию: существуют* w_1^*, $w_2^* \in [-1,1]$, *такие, что*

$$\begin{pmatrix} x_1^* - x_2^* \\ x_2^* - x_1^* \end{pmatrix} + \lambda \begin{pmatrix} w_1^* \left(A_1 x_1 + b_1\right) \\ w_2^* \left(A_2 x_2 + b_2\right) \end{pmatrix} = \mathbb{O}_{2n}.$$

5. Направление наискорейшего спуска функции Φ_λ

Предположим, что точка $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ такая, что $x_1 \neq x_2$ и $A_k x_k + b_k \neq 0$ при $k = 1, 2$. Ранее уже было установлено, что функция $\varphi(x)$ дифференцируема по направлениям. Поскольку функция $f(x)$ дифференцируемая, то штрафная функция $\Phi_\lambda(x) = f(x) + \lambda \varphi(x)$ дифференцируема по направлениям в точке x.

Далее будет показано, как можно находить направления наискорейшего спуска. Возможны следующие 4 случая:

1) $h_1(x_1) \neq 0$, $h_2(x_2) \neq 0$;

2) $h_1(x_1) = 0$, $h_2(x_2) \neq 0$;

3) $h_1(x_1) \neq 0$, $h_2(x_2) = 0$;

4) $h_1(x_1) = 0$, $h_2(x_2) = 0$.

Рассмотрим отдельно каждый из этих случаев.

Случай 1. Из (5) и (20) заключаем, что Φ_λ дифференцируема в точке x, причем ее градиент $G_\lambda(x) = \dfrac{\partial \Phi_\lambda(x)}{\partial x}$ равен

$$G_\lambda(x) = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \lambda \begin{pmatrix} \left(A_1 x_1 + b_1\right) \operatorname{sign} h_1(x_1) \\ \left(A_2 x_2 + b_2\right) \operatorname{sign} h_2(x_2) \end{pmatrix}. \tag{26}$$

Направление $g_\lambda(x) = -G_\lambda(x)/\|G_\lambda(x)\|$ является направлением наискорейшего спуска функционала Φ_λ в точке x.

Случай 2. Из (6) и (20) заключаем, что функция Φ_λ субдифференцируема в точке x, причем ее субдифференциал $\partial\Phi_\lambda(x)$ равен

$$\partial\Phi_\lambda(x) = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \\ + \lambda \operatorname{co}\left\{ \begin{pmatrix} -(A_1x_1 + b_1) \\ (A_2x_2 + b_2)\operatorname{sign} h_2(x_2) \end{pmatrix}, \begin{pmatrix} A_1x_1 + b_1 \\ (A_2x_2 + b_2)\operatorname{sign} h_2(x_2) \end{pmatrix} \right\}. \quad (27)$$

Каждый элемент $W \in \partial\Phi_\lambda(x)$ можно описать следующим образом:

$$W(\mu) = \begin{pmatrix} x_1 - x_2 + \mu\lambda\,(A_1x_1 + b_1) \\ x_2 - x_1 + \lambda\,(A_2x_2 + b_2)\operatorname{sign} h_2(x_2) \end{pmatrix},$$

где $\mu \in [-1, 1]$. Найдем минимальный по норме субградиент $W \in \partial\Phi_\lambda(x)$, т. е. решим задачу

$$\min_{W\in\partial\Phi_\lambda(x)} \|W\|^2 = \min_{\mu\in[-1,1]} \Big[\|x_1 - x_2 + \mu\lambda\,(A_1x_1 + b_1)\,\|^2 + \\ + \|x_2 - x_1 + \lambda\,(A_2x_2 + b_2)\operatorname{sign} h_2(x_2)\|^2 \Big].$$

Но прежде вычислим минимум на всей вещественной оси:

$$\min_{\mu\in\mathbb{R}} \Big[\|x_1 - x_2 + \mu\lambda\,(A_1x_1 + b_1)\,\|^2 + \|x_2 - x_1 + \lambda\,(A_2x_2 + b_2)\operatorname{sign} h_2(x_2)\|^2 \Big] = \\ = \min_{\mu\in\mathbb{R}} \Big[\mu^2\lambda^2\|A_1x_1 + b_1\|^2 + 2\mu\lambda(x_1 - x_2)^T\,(A_1x_1 + b_1) + \\ + \|x_1 - x_2\|^2 + \|x_2 - x_1 + \lambda\,(A_2x_2 + b_2)\operatorname{sign} h_2(x_2)\|^2 \Big].$$

При сделанных предположениях коэффициент при μ^2 положителен, поэтому этот минимум достигается в единственной точке

$$\mu^* = -\frac{(x_1 - x_2)^T\,(A_1x_1 + b_1)}{\lambda\|A_1x_1 + b_1\|^2}.$$

Напомним, что мы можем выбрать любое $\lambda > \lambda^*$ (см. **теорему 1**), поэтому при достаточно больших λ будет $\mu^* \in [-1, 1]$.

Вектор

$$W^* = \begin{pmatrix} x_1 - x_2 + \mu^*\lambda\,(A_1x_1 + b_1) \\ x_2 - x_1 + \lambda\,(A_2x_2 + b_2)\operatorname{sign} h_2(x_2) \end{pmatrix}$$

является наименьшим по норме субградиентом функции Φ_λ в точке x. Направление $g_\lambda(x) = -W^*/\|W^*\|$ является направлением наискорейшего спуска функционала Φ_λ в точке x.

Случай 3. Из (7) и (20) заключаем, что субдифференциал $\partial\Phi_\lambda(x)$ равен

$$\partial\Phi_\lambda(x) = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \\ + \lambda\,\mathrm{co}\left\{ \begin{pmatrix} (A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1) \\ -\,(A_2x_2 + b_2) \end{pmatrix}, \begin{pmatrix} (A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1) \\ A_2x_2 + b_2 \end{pmatrix} \right\}. \quad (28)$$

Каждый элемент $W \in \partial\Phi_\lambda(x)$ можно описать следующим образом:

$$W(\mu) = \begin{pmatrix} x_1 - x_2 + \lambda\,(A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1) \\ x_2 - x_1 + \mu\lambda\,(A_2x_2 + b_2) \end{pmatrix},$$

где $\mu \in [-1, 1]$. Найдем минимальный по норме субградиент $W \in \partial\Phi_\lambda(x)$, т. е. решим задачу

$$\min_{W \in \partial\Phi_\lambda(x)} \|W\|^2 = \min_{\mu \in [-1,1]} \Big[\|x_1 - x_2 + \lambda\,(A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1)\|^2 + \\ + \|x_2 - x_1 + \mu\lambda\,(A_2x_2 + b_2)\,\|^2 \Big].$$

Аналогично случаю 2 вначале решим задачу на всей вещественной оси:

$$\min_{\mu \in \mathbb{R}} \Big[\|x_1 - x_2 + \lambda\,(A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1)\|^2 + \|x_2 - x_1 + \mu\lambda\,(A_2x_2 + b_2)\,\|^2 \Big] = \\ = \min_{\mu \in \mathbb{R}} \Big[\mu^2\lambda^2\|A_2x_2 + b_2\|^2 + 2\mu\lambda(x_2 - x_1)^T\,(A_2x_2 + b_2) + \\ + \|x_2 - x_1\|^2 + \|x_1 - x_2 + \lambda\,(A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1)\|^2 \Big].$$

Здесь минимум достигается в единственной точке

$$\mu^* = -\frac{(x_2 - x_1)^T\,(A_2x_2 + b_2)}{\lambda\|A_2x_2 + b_2\|^2},$$

где при достаточно больших λ будет выполнено $\mu^* \in [-1, 1]$.

Тогда вектор

$$W^* = \begin{pmatrix} x_1 - x_2 + \lambda\,(A_1x_1 + b_1)\,\mathrm{sign}\,h_1(x_1) \\ x_2 - x_1 + \mu^*\lambda\,(A_2x_2 + b_2) \end{pmatrix}$$

является наименьшим по норме субградиентом функции Φ_λ в точке x. Направление $g_\lambda(x) = -W^*/\|W^*\|$ является направлением наискорейшего спуска функционала Φ_λ в точке x.

Наконец, рассмотрим случай $\varphi(x) = 0$.

Случай 4. Из (4) и (20) заключаем, что функция Φ_λ субдифференцируема в точке x, причем ее субдифференциал $\partial\Phi_\lambda(x)$ равен

$$\partial\Phi_\lambda(x) = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \lambda \operatorname{co} \left\{ \begin{pmatrix} w_1 \left(A_1 x_1 + b_1\right) \\ w_2 \left(A_2 x_2 + b_2\right) \end{pmatrix} \;\middle|\; w_k \in [-1, 1],\ k = 1, 2 \right\}. \quad (29)$$

Любой элемент $W \in \partial\Phi_\lambda(x)$ можно представить в следующем виде:

$$W(\mu_1, \mu_2) = \begin{pmatrix} x_1 - x_2 + \mu_1 \left(A_1 x_1 + b_1\right) \\ x_2 - x_1 + \mu_2 \left(A_2 x_2 + b_2\right) \end{pmatrix},$$

где $\mu_k \in [-\lambda, \lambda]$, $k = 1, 2$. Найдем минимальный по норме субградиент $W \in \partial\Phi_\lambda(x)$, т. е. решим задачу

$$\min_{W \in \partial\Phi_\lambda(x)} \|W\|^2 = \min_{\substack{\mu_k \in [-\lambda, \lambda] \\ k=1,2}} \Big[\|x_1 - x_2 + \mu_1 \left(A_1 x_1 + b_1\right)\|^2 + \\ + \|x_2 - x_1 + \mu_2 \left(A_2 x_2 + b_2\right)\|^2 \Big].$$

И вновь вначале найдем

$$\min_{\substack{\mu_k \in \mathbb{R} \\ k=1,2}} \Big[\|x_1 - x_2 + \mu_1 \left(A_1 x_1 + b_1\right)\|^2 + \|x_2 - x_1 + \mu_2 \left(A_2 x_2 + b_2\right)\|^2 \Big] =$$

$$= \min_{\substack{\mu_k \in \mathbb{R} \\ k=1,2}} \Big[\mu_1^2 \|A_1 x_1 + b_1\|^2 + \mu_2^2 \|A_2 x_2 + b_2\|^2 + 2\mu_1 (x_1 - x_2)^T \left(A_1 x_1 + b_1\right) + \\ + 2\mu_2 (x_2 - x_1)^T \left(A_2 x_2 + b_2\right) + 2\|x_1 - x_2\|^2 \Big].$$

Приравнивая нулю производные по μ_1 и μ_2, получим систему линейных уравнений

$$\begin{pmatrix} \|A_1 x_1 + b_1\|^2 & 0 \\ 0 & \|A_2 x_2 + b_2\|^2 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} (x_2 - x_1)^T \left(A_1 x_1 + b_1\right) \\ (x_1 - x_2)^T \left(A_2 x_2 + b_2\right) \end{pmatrix}.$$

При сделанных предположениях определитель этой системы положителен, поэтому она имеет единственное решение

$$\begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} = \begin{pmatrix} \dfrac{(x_2 - x_1)^T \left(A_1 x_1 + b_1\right)}{\|A_1 x_1 + b_1\|^2} \\ \dfrac{(x_1 - x_2)^T \left(A_2 x_2 + b_2\right)}{\|A_2 x_2 + b_2\|^2} \end{pmatrix}.$$

Вектор

$$W^* = \begin{pmatrix} x_1 - x_2 + \mu_1^* \left(A_1 x_1 + b_1 \right) \\ x_2 - x_1 + \mu_2^* \left(A_2 x_2 + b_2 \right) \end{pmatrix} \tag{30}$$

является наименьшим по норме субградиентом функции Φ_λ в точке x и не зависит от параметра λ.

Если $\|W^*\| > 0$, то направление $g(x) = -W^*/\|W^*\|$ является направлением наискорейшего спуска функционала Φ_λ в точке x.

Таким образом, используя классическую вариацию (2), имеем

$$\Phi_\lambda(x_\varepsilon) = \Phi_\lambda(x) + \varepsilon H_\lambda(x, v) + o(\varepsilon). \tag{31}$$

Здесь функцию $H_\lambda(x, v)$ можно представить в виде

$$H_\lambda(x, v) = \max_{W \in \partial\Phi_\lambda(x)} \langle v, W \rangle,$$

где субдифференциал $\partial\Phi_\lambda(x)$, в зависимости от случая, принимает вид (26)–(29).

6. Метод наискорейшего спуска

Пусть $x_* \in X$ — точка минимума функционала $\Phi_\lambda(x)$ на $\mathbb{R}^{2n}$, в частности, $\varphi(x_*) = 0$. Как следует из **теоремы 5**, необходимое условие (25) эквивалентно существованию w_1^*, $w_2^* \in [-1, 1]$, для которых

$$\begin{pmatrix} x_1^* - x_2^* \\ x_2^* - x_1^* \end{pmatrix} + \lambda \begin{pmatrix} w_1^* \left(A_1 x_1 + b_1 \right) \\ w_2^* \left(A_2 x_2 + b_2 \right) \end{pmatrix} = \mathbb{O}_{2n}.$$

Это условие, в свою очередь, эквивалентно тому, что

$$\mathbb{O}_{2n} \in \partial\Phi_\lambda(x_*). \tag{32}$$

Точка x_*, в которой выполнено условие (32), называется стационарной.

Итак, если точка $x \in \mathbb{R}^{2n}$ не является стационарной точкой функционала Φ_λ, то можно найти направление наискорейшего спуска функционала Φ_λ в точке x. Опишем метод наискорейшего спуска для нахождения стационарных точек.

Выберем произвольное $z^0 \in \mathbb{R}^{2n}$. Пусть уже найдено $z^k \in \mathbb{R}^{2n}$. Если $\varphi(z^k) = 0$ и выполнено условие (32), то точка z^k является стационарной, и процесс прекращается.

Если же $\varphi(z^k) \neq 0$ или $\varphi(z^k) = 0$, но условие (32) не выполнено, то возьмем функцию $G_{k\lambda} = G_\lambda(z^k)$ — наименьший по норме субградиент функционала Φ_λ в точке z^k.

Далее решается задача одномерной минимизации

$$\min_{\beta \geqslant 0} \Phi_\lambda(z^k - \beta G_{k\lambda}) = \Phi_\lambda(z^k - \beta_k G_{k\lambda}). \tag{33}$$

Теперь положим $z^{k+1} = z^k - \beta_k G_{k\lambda}$. Имеем $\Phi_\lambda(z^{k+1}) < \Phi_\lambda(z^k)$.

Заметим, что величина шага спуска β в задаче (33) вычисляется аналитически.

К сожалению, описанный процесс может и не привести к стационарной точке, поскольку субдифференциальное отображение $\partial\Phi_\lambda(x)$ не является непрерывным, как функция x в метрике Хаусдорфа [6, 9].

7. Метод гиподифференциального спуска

Вместо разложения (31) для вариации (2) можно получить представление «гиподифференциального» типа, а именно

$$\Phi_\lambda(x_\varepsilon) = \Phi_\lambda(x) + H_\lambda(\varepsilon, x, v) + o(\varepsilon), \tag{34}$$

где

$$\begin{aligned} H_\lambda(\varepsilon, x, v) &= \varepsilon v^T \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix} + \\ &+ \lambda\Big[\max\big\{ h_1(x_1) - |h_1(x_1)| + \varepsilon v_1^T (A_1 x_1 + b_1), \\ &\qquad - h_1(x_1) - |h_1(x_1)| - \varepsilon v_1^T (A_1 x_1 + b_1) \big\} + \\ &+ \max\big\{ h_2(x_2) - |h_2(x_2)| + \varepsilon v_2^T (A_2 x_2 + b_2), \\ &\qquad - h_2(x_2) - |h_2(x_2)| - \varepsilon v_2^T (A_2 x_2 + b_2) \big\} \Big]. \end{aligned}$$

Из (34) следует, что функционал $\Phi_\lambda(x)$ гиподифференцируем в точке x [9], и его гиподифференциал $d\Phi_\lambda(x)$ равен

$$\begin{aligned} d\Phi_\lambda(x) &= \left\{ \begin{pmatrix} 0 \\ Q(x) \end{pmatrix} \right\} + \\ &+ \lambda\Bigg(\mathrm{co}\left\{ \begin{pmatrix} h_1(x_1) - |h_1(x_1)| \\ h'_{1z}(x_1) \end{pmatrix}, \begin{pmatrix} -h_1(x_1) - |h_1(x_1)| \\ -h'_{1z}(x_1) \end{pmatrix} \right\} + \\ &\qquad + \mathrm{co}\left\{ \begin{pmatrix} h_2(x_2) - |h_2(x_2)| \\ h'_{2z}(x_2) \end{pmatrix}, \begin{pmatrix} -h_2(x_2) - |h_2(x_2)| \\ -h'_{2z}(x_2) \end{pmatrix} \right\} \Bigg), \end{aligned}$$

где

$$Q(x) = \begin{pmatrix} x_1 - x_2 \\ -(x_1 - x_2) \end{pmatrix}, \quad h'_{1z}(x_1) = \begin{pmatrix} A_1 x_1 + b_1 \\ \mathbb{O}_n \end{pmatrix}, \quad h'_{2z}(x_2) = \begin{pmatrix} \mathbb{O}_n \\ A_2 x_2 + b_2 \end{pmatrix}.$$

Отметим, что отображение $d\Phi_\lambda(x)$ является непрерывным в метрике Хаусдорфа [9]. Нетрудно также показать, что необходимое условие минимума функции $\Phi_\lambda(x)$ (см. (32)) эквивалентно условию

$$[0, \mathbb{O}_{2n}] \in d\Phi_\lambda(x_*). \tag{35}$$

Перед тем, как приступить к поиску минимального по норме гипоградиента $W \in \underline{d}\Phi_\lambda(x)$ заметим, что гиподифференциал $\underline{d}\Phi_\lambda(x)$ можно представить в следующем виде $\underline{d}\Phi_\lambda(x) = \big\{ W(\mu_1, \mu_2) \;\big|\; |\mu_1| \leqslant \lambda,\ |\mu_2| \leqslant \lambda \big\}$, где

$$W(\mu_1, \mu_2) = \begin{pmatrix} \mu_1 h_1(x_1) + \mu_2 h_2(x_2) - \lambda\varphi(x) \\ \mu_1 h'_{1z}(x_1) + \mu_2 h'_{2z}(x_2) + Q(x) \end{pmatrix}.$$

Найдем

$$\min_{W \in d\Phi_\lambda(x)} \|W\|^2 = \min_{\substack{\mu_k \in [-\lambda, \lambda] \\ k=1,2}} \|W(\mu_1, \mu_2)\|^2 = \|W^*\|^2.$$

Однако вначале решим следующую задачу

$$\min_{\substack{\mu_k \in \mathbb{R} \\ k=1,2}} L(\mu_1, \mu_2),$$

где

$$L(\mu_1, \mu_2) := \|W(\mu_1, \mu_2)\|^2 =$$
$$= \big[\mu_1 h_1(x_1) + \mu_2 h_2(x_2) - \lambda\varphi(x)\big]^2 + \|\mu_1 h'_{1z}(x_1) + \mu_2 h'_{2z}(x_2) + Q(x)\|^2.$$

Приравняем к нулю производные функции $L(\mu_1, \mu_2)$ по μ_1, μ_2. Получим систему

$$D\mu = \eta, \tag{36}$$

где

$$D = \begin{pmatrix} h_1^2(x_1) + (h'_{1z}(x_1))^T h'_{1z}(x_1) & h_1(x_1) h_2(x_2) \\ h_1(x_1) h_2(x_2) & h_2^2(x_2) + (h'_{2z}(x_2))^T h'_{2z}(x_2) \end{pmatrix},$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \eta = \begin{pmatrix} \lambda\varphi(x) h_1(x_1) - (h'_{1z}(x_1))^T Q(x) \\ \lambda\varphi(x) h_2(x_2) - (h'_{2z}(x_2))^T Q(x) \end{pmatrix}.$$

Вновь предположим, что точка $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ из $\mathbb{R}^{2n}$ такая, что $x_1 \neq x_2$ и $A_k x_k + b_k \neq 0$ при $k = 1, 2$. Тогда $\det D > 0$, и система (36) имеет единственное решение

$$\mu^* = D^{-1}\eta.$$

При достаточно больших λ будет $|\mu_k^*| \leqslant \lambda$, $k = 1, 2$. Следовательно, при таких значениях параметра λ имеем

$$W^* = \begin{pmatrix} \mu_1^* h_1(x_1) + \mu_2^* h_2(x_2) - \lambda\varphi(x) \\ \mu_1^* h'_{1z}(x_1) + \mu_2^* h'_{2z}(x_2) + Q(x) \end{pmatrix}.$$

Вектор

$$G^*(x) = \mu_1^* h'_1(x_1) + \mu_2^* h'_2(x_2) + Q(x) \tag{37}$$

является минимальным по норме гипоградиентом функционала $\Phi_\lambda(x)$. Если $\|G^*\| > 0$, т. е. точка x не является стационарной, то направление $g(x) = -\dfrac{G^*(x)}{\|G^*(x)\|}$ является направлением спуска функционала Φ_λ в точке x. В отличие от направления наискорейшего спуска, направление $g(x)$, как функция x, является непрерывным.

Таким образом, если точка $x \in \mathbb{R}^{2n}$ не является стационарной точкой функционала Φ_λ, то можно найти направление спуска функционала Φ_λ в точке x. Опишем следующий метод гиподифференциального спуска для нахождения стационарных точек, т. е. точек удовлетворяющих условию (32) или (35).

Выберем произвольное $z_0 \in \mathbb{R}^{2n}$. Пусть уже найдено $z_k \in \mathbb{R}^{2n}$. Если $\varphi(z_k) = 0$ и выполнено условие (35), то точка z_k является стационарной, и процесс прекращается. Если же условие (35) не выполнено, то возьмем вектор $G^*(z_k)$ — минимальный по норме гипоградиент функции Φ_λ в точке z_k. Теперь найдем

$$\min_{\beta \geqslant 0} \Phi_\lambda(z_k - \beta G^*(z_k)) = \Phi_\lambda(z_k - \beta_k G^*(z_k)) \tag{38}$$

и положим $z_{k+1} = z_k - \beta_k G^*(z_k)$. Имеем $\Phi_\lambda(z_{k+1}) < \Phi_\lambda(z_k)$. Пользуясь непрерывностью в метрике Хаусдорфа гиподифференциального отображения как функции z, можно доказать, что описанный метод сходится в следующем смысле: $\|G^*(z_k)\| \longrightarrow 0$.

8. Двухшаговый гипоградиентный метод

Пусть z_0 — некоторое начальное приближение из $\mathbb{R}^{2n}$. Будем строить последовательность $\{z_k\}$ по формулам

$$z_{k+1} = z_k - \beta_k W_k, \quad k = 0, 1, \ldots, \tag{39}$$

где

$$W_0 = G(z_0), \quad W_k = G(z_k) + \gamma_k W_{k-1}, \quad k = 1, 2, \ldots.$$

Здесь $G(z) = \mu_1^* h'_{1z}(x_1) + \mu_2^* h'_{2z}(x_2) + Q(z)$ — минимальный по норме гипоградиент функционала $\Phi_\lambda(z)$. Величина β_k определяется (аналитически) так же, как и в методе гиподифференциального спуска, т. е.

$$\min_{\beta \geqslant 0} \Phi_\lambda(z_k - \beta W_k) = \Phi_\lambda(z_k - \beta_k W_k),$$

а γ_k — по формуле Полака–Райбера

$$\gamma_k = \frac{\langle G(z_k), G(z_k) - G(z_{k-1})\rangle}{\langle G(z_{k-1}), G(z_{k-1})\rangle}. \tag{40}$$

9. Численные эксперименты

Были проведены массовые вычисления по сравнению эффективности метода гиподифференциального спуска и двухшагового гипоградиентного метода в математическом пакете MATLAB. Брались m пар эллипсоидов в n-мерном евклидовом пространстве при n, равном 2, 3, 50 и при m, равном 10, 50 и 100. Эллипсоиды формировались с помощью функции генерирования псевдослучайных чисел. В приведенных ниже таблицах указано время T (в сек.). Все вычисления проводились с точностью 10^{-5}.

Таблица 1. Метод гиподифференциального спуска.

n \ m	10	50	100
2	8,526	61,275	161,099
3	16,160	93,701	301,014
50	37,291	310,324	405,597

Таблица 2. Двухшаговый гипоградиентный метод.

n \ m	10	50	100
2	7,579	58,163	130,448
3	14,450	87,375	185,045
50	24,481	199,447	358,630

Заключение

В работе описаны три алгоритма решения задачи о нахождении расстояния между двумя эллипсоидами. Разработано программное обеспечение в системах MATLAB и Maple, реализующее данные алгоритмы. Из проведенных экспериментов следует, что методы гиподифференциального спуска и двухшаговый гипоградиентный более универсальные, чем метод «шаров» [3], и более эффективные нежели методы [2, 4].

ЛИТЕРАТУРА

1. Утешев А. Ю. *Вычисление расстояний между геометрическими объектами* [Электронный ресурс] // Записная книжка профессора Утешева [сайт]. [2007]. http://pmpu.ru/vf4/algebra2/optimiz/distance

2. Утешев А. Ю., Яшина М. В. *Нахождение расстояния от эллипсоида до плоскости и квадрики в* $\mathbb{R}^n$ // Доклады АН, 2008. Т. 419, № 4. С. 471–474.

3. Lin A., Han S. P. *On the distance between two ellipsoids* // SIAM Journal on Optimization, 2002. Vol. 13. P. 298–308.

4. Лебедев Д. М., Полякова Л. Н, *Задача проектирования нулевой точки на квадрику* // Вестник СПбГУ. Сер. 10, 2013. Вып. 1. С. 11–17.

5. Тамасян Г.Ш., Чумаков А.А. *Нахождение расстояния между эллипсоидами* // Дискрет. анализ и исслед. операций. 2014. Т. 21. № 3. С. 87-102.

6. Демьянов В. Ф. *Условия экстремума и вариационное исчисление* М.: Высшая школа, 2005. 335 с.

7. Еремин И. И. *Метод «штрафов» в выпуклом программировании* // Доклады АН СССР, 1967. Т. 143, № 4. С. 748–751.

8. Долгополик М. В. *Точные штрафные функции в негладкой оптимизации* // Семинар «CNSA & NDO». Избранные доклады. 8 мая 2014 г. (`http://www.apmath.spbu.ru/cnsa/reps14.shtml#0508`)

9. Демьянов В.Ф. Рубинов А.М. *Основы негладкого анализа и квазидифференциальное исчисление.* М.: Наука, 1990. 432 с.