
 ЗАМЕТКИ
И ПИСЬМА

ОПТИМИЗАЦИЯ ПОЛЯ ДЕЙСТВИЯ КОНЕЧНОГО ЧИСЛА ИСТОЧНИКОВ В НЕПРЕРЫВНОЙ СРЕДЕ

© 2009 г. Л. Я. Клеппер

(Москва)

На языке экономических моделей в работе рассматривается задача оптимального размещения объектов инфраструктуры, создающих в совокупности территориальное “поле обслуживания” для “клиентов”, находящихся в данном ареале. К настоящему времени существует значительное число эвристических методов решения подобных задач, основанных на тех или иных представлениях о характере взаимодействия размещаемых объектов и клиентов; при этом описание взаимодействия часто дается по аналогии с физическими законами или их модификациями. Так, для приближенного описания транспортных потоков использовались гравитационные и энтропийные модели (Браиловский, 1979). При решении проблемы создания новых производственных комплексов в работах (Фаерман, 1971; Фаерман, Клеппер, Вейцман, 1979) был разработан метод общего геопотенциала, который позволяет оптимизировать районы размещения комплексов. Метод основан на аналогии с физическим взаимодействием электрических зарядов.

С содержательной точки зрения в экономических моделях размещаемые объекты представляют собой “пункты обслуживания”, такие как склады, магазины, ремонтные мастерские и т.п., и задача оптимального размещения состоит в том, чтобы максимально приблизить эти пункты к клиентам. Для автора данной статьи такого рода задачи представляют интерес в связи с его работами в области оптимального планирования лучевой терапии злокачественных опухолей¹. В статье рассматриваемые задачи интерпретируются в указанном контексте: в качестве поля обслуживания выступает пораженная область, все точки которого являются “клиентами”, а пунктами обслуживания выступают источники облучения, помещаемые внутри пораженной области и создающие терапевтическое радиационное поле, подавляющее очаги поражения. Проблема состоит в том, чтобы распределить заданное число источников облучения наилучшим (в определенном смысле) образом.

ПОСТАНОВКА ЗАДАЧИ В ОБЩЕЙ ФОРМЕ

Пусть G — ограниченная область обслуживания в k -мерном пространстве R , точки которой обозначаются y ; X — конечное множество точечных источников x (реальные значения $k = 1, 2, 3$). Предполагается, что область G выпуклая и X является подмножеством в G , т.е. источники располагаются внутри обслуживаемой области. Для удобства будем называть множество X — “Источником”.

Воздействие каждого источника на среду характеризуется *функцией действия* $d(r)$, зависящей от расстояния $r > 0$ между источником x и точкой среды $y \in G$. Относительно этой функции предполагается, что она обладает следующими свойствами.

Свойство 1. Функция $d(r)$ непрерывная и строго убывает по r (т.е. $d'(r) < 0$), причем $d(+0) = +\infty$, $d(\infty) = 0$.

Свойство 2. Функция d выпуклая, т.е. $d''(r) > 0$.

Свойство 3. Функция d однородная отрицательной степени; это означает, что при некотором $\theta > 0$ выполняется соотношение $d(\lambda r) = \lambda^{-\theta} d(r)$, $r > 0 \quad \forall \lambda > 0$.

¹ В ЦЭМИ АН СССР в 1970-е годы зародилось новое научное направление в лучевой терапии злокачественных опухолей — разработка методов оптимального дозирования. Оно было развито в большом цикле работ с участием автора (около 30 публикаций). В монографии (Swan, 1981) цикл этих исследований получил высокую оценку.

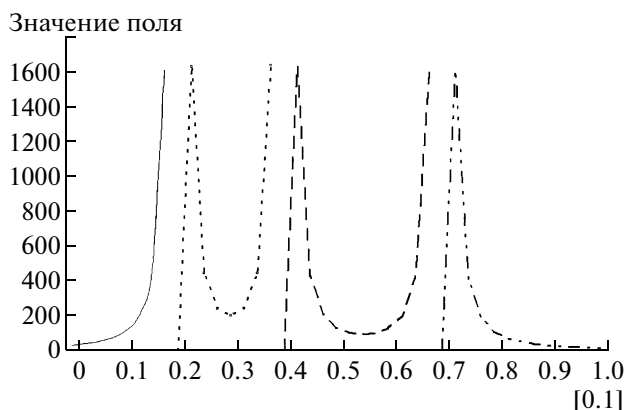


Рис. 1. Поле действия, образованное тремя источниками с четырьмя точками минимума.

Перечисленными свойствами, при дополнительном (не ограничивающем общности) условии нормировки $d(1) = 1$, обладает только степенная функция

$$d(r) = 1 / r^{\theta}, \theta > 0; \quad (1)$$

типичные случаи $\theta = 2$ и $\theta = 1$.

Совокупный эффект действия Источника X образует *поле действия* $D(X)$. По определению, значение поля в точке $y \in G$ задается выражением

$$D(X, y) := \sum_{x \in X} d(|x - y|), y \notin X, \quad (2)$$

где $|x - y|$ — расстояние между точками x и y .

В общем виде рассматриваемая задача формулируется следующим образом: для заданного числа источников n разместить множество X таким образом, чтобы максимизировать минимальный уровень поля $D(X)$ (предполагается, что “слипание” источников недопустимо), т.е.

$$m(X) := \min_{y \in G} D(X, y) \rightarrow \max_X, (X \subset G, |X| = n); \quad (3)$$

максимальное значение критерия обозначим через M .

Примечание. Во внутритканевой лучевой терапии для достижения полного терапевтического эффекта (безрецидивного излечения опухолевого заболевания) необходимо так распределить источники радиоактивного излучения в опухоли, чтобы дозовое поле $D(X)$ было как можно более однородным. Именно это требование однородности и заложено в критерии (3). Дело в том, что в области низких уровней облучения (в области локальных минимумов поля действия) возможен рецидив заболевания и, наоборот, в области высоких доз могут возникнуть трудноизлечимые лучевые некрозы.

Замечание. Из свойства 3 вытекает, что структура оптимального размещения зависит только от формы, а не от размера области G : при ее гомотетии (сжатии или расширении подобным образом), оптимальное размещение X подвергается той же гомотетии.

ОДНОМЕРНЫЙ СЛУЧАЙ: РАЗМЕЩЕНИЕ ИСТОЧНИКОВ НА ОТРЕЗКЕ

В случае, когда область обслуживания G является отрезком вещественной прямой R , задача имеет наиболее простое решение. Этот случай интересен и сам по себе, и как “мостик” для перехода к двумерному и трехмерному случаям.

1. Предварительный анализ. В одномерном случае ключевыми являются следующие соображения.

1. При оптимальном размещении источников ни один из них не будет находиться на конце отрезка G , так как в противном случае концевой источник можно слегка сместить внутрь отрезка и увеличить тем самым величину $m(X)$. Размещения, удовлетворяющие указанному условию, назовем *допустимыми*. При общем числе источников n при всяком допустимом размещении отрезок G будет разбит на $n + 1$ интервалов разбиения, из которых $n - 1$ будут внутренними, а два — концевыми (рис. 1).

2. На любом интервале разбиения Δ поле действия $D(X)$ является выпуклой функцией (так как в силу свойства 2 она есть сумма выпуклых функций); следовательно, на Δ точка минимума существует и является единственной. Для концевых интервалов точка минимума будет концом отрезка G , для внутренних интервалов она будет находиться строго внутри интервала (так как на концах интервала значение поля равно $+\infty$). С учетом соображения 1 всякое допустимое размещение имеет $n + 1$ локальный минимум.

3. Рассмотрим (при некотором размещении источников) какой-либо из интервалов разбиения Δ . Обозначим минимальный уровень поля действия на этом интервале через $m_{\text{соб}}$ (собственный минимум), а вне Δ — через $m_{\text{чуж}}$ (чужой минимум).

Лемма 1. Если при размещении X имеем, что $m_{\text{соб}} \neq m_{\text{чуж}}$, то это размещение можно улучшить — повысить глобальный минимум $m(X)$.

Доказательство. Пусть $m_{\text{соб}} > m_{\text{чуж}}$. Тогда глобальный минимум $m(X)$ находится вне интервала Δ и равен $m_{\text{чуж}}$. Возможны два случая — интервал Δ концевой или интервал Δ внутренний. Расширим интервал Δ на некоторую величину, причем в первом случае раздвинем только внутренний конец отрезка, а во втором случае — оба (внутренних) конца. Это позволит нам понизить уровень $m_{\text{соб}}$ и одновременно повысить уровень $m_{\text{чуж}}$ (в этом легко убедиться), и тем самым повысим глобальный минимум.

Пусть теперь $m_{\text{соб}} < m_{\text{чуж}}$; тогда глобальный минимум совпадает с $m_{\text{соб}}$. В этой ситуации будем сужать интервал Δ по правилу, указанному выше. Эффект будет такой же, как и в предыдущем случае — глобальный минимум повысится.

Из леммы 1 вытекают два важных следствия.

Предложение 1 (условие оптимальности размещения X). Для того чтобы допустимое размещение X было оптимальным, необходимо, чтобы каждый локальный минимум имел один и тот же уровень. В этом случае уровень будет глобальным и равен $m(X)$.

Доказательство. Если не все локальные минимумы равны, то существует по крайней мере один интервал Δ , уровень которого $m_{\text{соб}}$ отличается от $m(X)$. Согласно лемме 1, такое размещение X не оптимально.

Лемма 1 дает конструктивный алгоритм построения размещения, все локальные минимумы которого одинаковы (такое размещение назовем *сбалансированным* или, коротко, *сб-размещением*); тем самым доказано существование сб-размещения.

Предложение 2. Если размещение построено согласно указанному в лемме 1 алгоритму, то существует *сбалансированное* размещение (сб-размещение).

Докажем теперь, что условие сбалансированности является не только необходимым, но и достаточным для оптимальности размещения. Для этого достаточно показать, что сб-размещение единственно.

2. Единственность сбалансированного размещения. Доказательство единственности сб-размещения требует рассмотрения общей задачи, которую назовем *возмущенной*. Она отличается от первоначальной задачи тем, что поле действия формируется не на “чистом листе”, а на некотором заданном *фоновом поле*; т.е. вместо (2) рассматривается поле

$$D_F(X, y) := F(y) + \sum_{x \in X} d(x - y), \quad y \in G, \quad y \notin X, \quad (4)$$

где F — заданная непрерывная функция на G . Ниже в процессе доказательства область G и функция F могут меняться; поэтому для большей четкости рассуждений исходные данные обозначим через G_0 и F_0 .

Предложение 3. Для произвольного отрезка G_0 и заданной на нем непрерывной функции F_0 сбалансированное размещение возмущенного поля (4) существует и единственно.

Доказательство. Проведем доказательство в два этапа.

1. К возмущенному полю применима лемма 1, так как наличие фона не препятствует ходу рассуждений, использованному при ее доказательстве, то оно полностью переносится на возмущенное поле. Как и раньше, из леммы 1 следует существование сб-размещения.

2. Единственность сб-размещения докажем методом математической индукции по числу источников. Пусть $n = 1$; тогда при любом допустимом размещении источника существуют два интервала разбиения. Проследим, как меняются их собственные минимальные уровни при перемещении источника от левого конца отрезка G_0 к правому концу. Левый уровень $m_{\text{соб}}$ строго убывает от $+\infty$ до некоторого конечного значения (зависящего от функции F_0); правый уровень $m_{\text{соб}}$ строго возрастает от начального значения до $+\infty$. Соответствующие кривые (графики этих уровней) обязательно пересекаются, причем в единственной точке. Тем самым для одного источника предложение доказано.

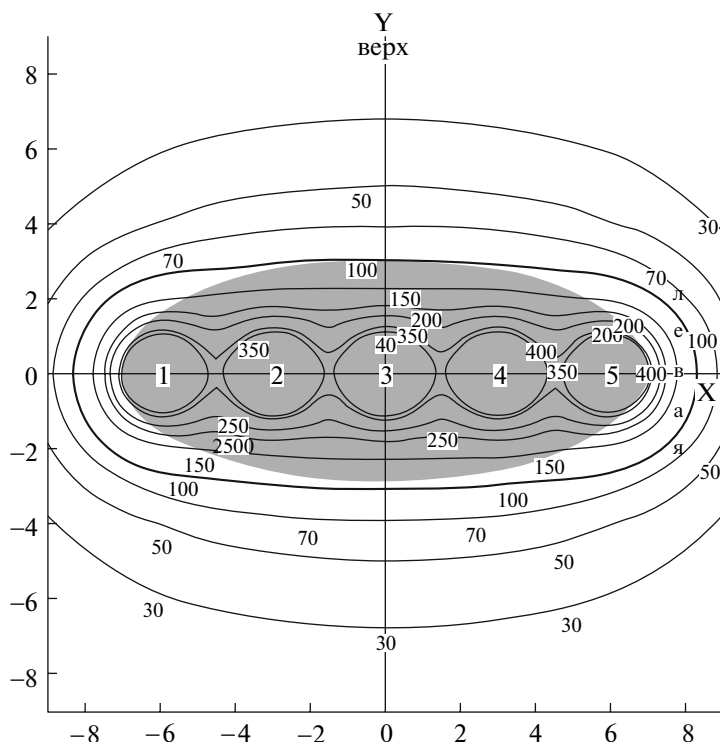


Рис. 2. Оптимальное размещение источников в вытянутой эллиптической области, которая мало отличается от прямолинейного отрезка. Распределение источников такое же, как и на отрезке, но только каждый внутренний локальный минимум исчезает, превращаясь в седловую точку.

Перейдем теперь к общему индуктивному шагу. Предположим, что единственность сб-размещения доказана для $n - 1$ ($n \geq 2$) источника, причем для произвольных отрезков G и непрерывной функции F на нем. Докажем утверждение для n источников и заданных $G_0 = [a, b]$ и F_0 . Поместим первый (самый левый) источник в некоторую точку x отрезка $[a, b]$ и построим сб-размещение остальных $n - 1$ источников на отрезке $G = [x, b]$ с фоновой функцией на нем, образующейся из первоначального фона F_0 и источника x , т.е. с функцией $F_x(y) := F_0(y) + d(x - y)$, $y \in (x, b]$.

Согласно индуктивному предположению, такое размещение существует и единственно; полученный сб-уровень обозначим через $M(x)$. Важно, что функция $M(x)$ строго возрастает по $x \in [a, b]$ (так как функция F_x строго возрастает по x и, кроме того, с ростом x сужается область минимизации (отрезок $[x, b]$) — от некоторого начального значения до $+\infty$). В то же время собственный уровень $m_{\text{соб}}$ левого интервала $[a, x]$ строго убывает по x от $+\infty$ до некоторого конечного уровня. Соответствующие кривые обязательно пересекаются в единственной точке, которая и определяет сб-размещение для всех n источников. Предложение доказано.

3. Итоговый результат. Объединяя предложения 1 и 3, получаем теорему.

Теорема. В случае, когда обслуживаемая область является прямолинейным отрезком, оптимальное размещение источников достигается тогда и только тогда, когда все локальные минимумы совокупного поля действия равны между собой. Это размещение существует и единственно (следовательно, оно симметрично относительно центра отрезка).

ДВУМЕРНЫЙ СЛУЧАЙ: ДВУХТОЧЕЧНЫЙ ИСТОЧНИК В ЭЛЛИПТИЧЕСКОЙ ОБЛАСТИ

Особенностью, принципиально отличающей двумерный случай от одномерного, является появление седловых точек. Этот эффект хорошо иллюстрируется на примере сильно вытянутой области, мало отличающейся от прямолинейного отрезка; оптимальное размещение практически останется тем же, что на отрезке, но только каждый *внутренний локальный минимум исчезает, превращаясь в седловую точку*, и заменяется на два крайних минимума (рис. 2). При любой выпук-

лой форме области G (не обязательно даже эллиптической) все локальные минимумы поля действия двухточечного Источника (их всегда ровно четыре) располагаются на границе поля обслуживания.

1. Двухточечный Источник с функцией действия $d(r) = 1 / r^2$. Этот пример очень удобен для иллюстрации, так как его можно полностью решить аналитически.

Пусть a, b : $a \geq b$, — полуоси эллипса с центром в начале координат плоскости (x, y) . Ясно, что при оптимальном размещении двух источников они будут располагаться симметрично на большой оси; пусть их абсциссы обозначаются x ($x > 0$) и $-x$. Условием оптимальности является равенство поля действия в вершинах эллипса, т.е.

$$\frac{1}{(a+x)^2} + \frac{1}{(a-x)^2} = 2 \frac{1}{b^2 + x^2}. \quad (5)$$

После преобразований равенство (5) приводит к квадратному уравнению, из которого получаем

$$x = a f(\lambda), \quad f(\lambda) := \sqrt{(1-\lambda)/(3+\lambda)}, \quad \lambda = (b/a)^2, \quad (6)$$

при этом

$$\mu = \mu(\lambda) = D(A) = D(B) = \left(\frac{1}{a^2}\right) \frac{2}{\lambda + f^3(\lambda)} = \frac{2(3+\lambda)}{a^2(1+\lambda)}.$$

Функция $f(\lambda)$ и величина минимума $M(\lambda)$ убывают по $\lambda \in [0, 1]$. Значению $\lambda = 0$ соответствует одномерный отрезок, здесь $x = a / \sqrt{3} = 0.577a$, $\mu = 6 / a^2$; значению $\lambda = 1$ отвечает круг, здесь $x = 0$, т.е. оптимальное размещение двух источников в круге — это поместить их оба в центр круга. Таким образом, для круга задача с двумя источниками вырождается в случай с одним источником.

2. Рассмотрим тот же пример, но с функцией действия $d(r) = 1 / r$. Рассуждая аналогично, получаем

$$\frac{1}{a+x} + \frac{1}{a-x} = 2 \frac{1}{\sqrt{b^2 + x^2}}. \quad (7)$$

Это уравнение похоже на уравнение (5), но в отличие от него уравнение (7) приводится к биквадратному, и если обозначить через $t := (x / a)^2$, то получим квадратное уравнение относительно t :

$$\lambda + t = (1-t)^2 \rightarrow t = (3 - \sqrt{5+4\lambda}) / 2$$

(здесь λ — то же, что в (6), т.е. $x = a \times g(\lambda)$ $g(\lambda) := \sqrt{3 - 0.5\sqrt{5+4\lambda}}$).

Как и в предыдущем случае, $g(\lambda)$ убывает по λ , при этом для отрезка $\lambda = 0$, $g(\lambda) = 0.618$, для круга $\lambda = 1$, $g(1) = 0$ (заключение относительно круга остается прежним). Нетрудно убедиться, что

$$g(\lambda) > f(\lambda) \quad \forall \lambda \in [0, 1). \quad (8)$$

4. ДВУМЕРНЫЙ СЛУЧАЙ: АЛГОРИТМ РАЗМЕЩЕНИЯ ИСТОЧНИКОВ В ПЛОСКОЙ ОБЛАСТИ ПРОИЗВОЛЬНОЙ ФОРМЫ

Для плоской области произвольной формы задача (3) не может быть решена аналитически. Для численного решения задачи необходимо организовать итерационный процесс. Предлагаемый ниже эвристический алгоритм итерационного процесса основан на той же идее, которая использовалась при доказательстве леммы 1. А именно, если при некотором размещении источников X точка y является точкой глобального минимума ($D(X, y) = m(X)$) и $x \in X$ — ближайший (или один из ближайших) к точке y источников, то, смещая источник x в сторону точки y (по радиальному направлению xy) на некоторую достаточно малую величину (расстояние) l , мы заведомо повысим значение минимума $m(X)$.

1. Алгоритм итерационного процесса.

Шаг 0. Покрываем область обслуживания G дискретной сеткой узлов с достаточно мелким шагом h (делаем задачу дискретной). В дальнейшем считаем, что узлы сетки представляют истинную область обслуживания G (т.е. будем вычислять поле обслуживания только в узловых точках).

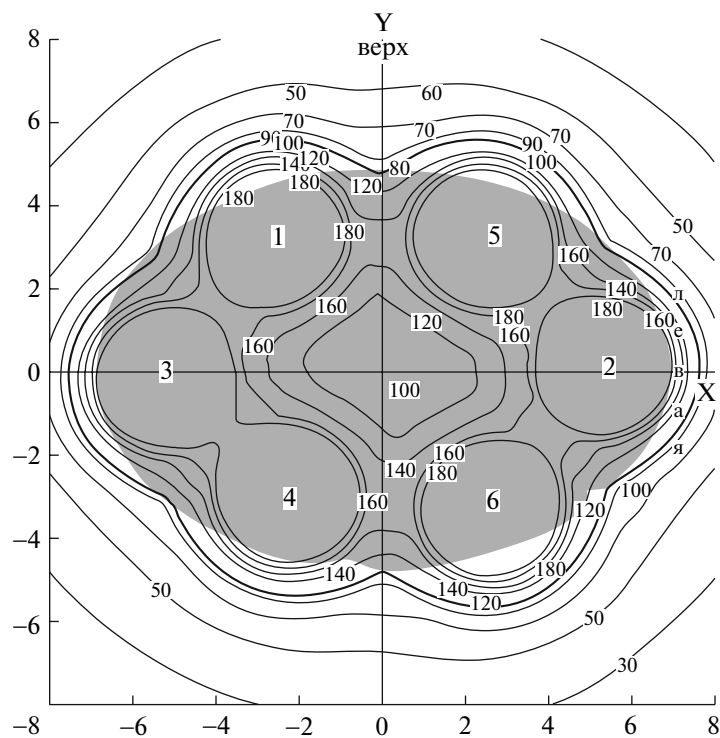


Рис. 3. Оптимальное размещение шести источников в эллипсе с отношением осей 5:7. За 100% принята изолиния поля действия, отвечающая глобальному минимуму (выделена полужирной линией).

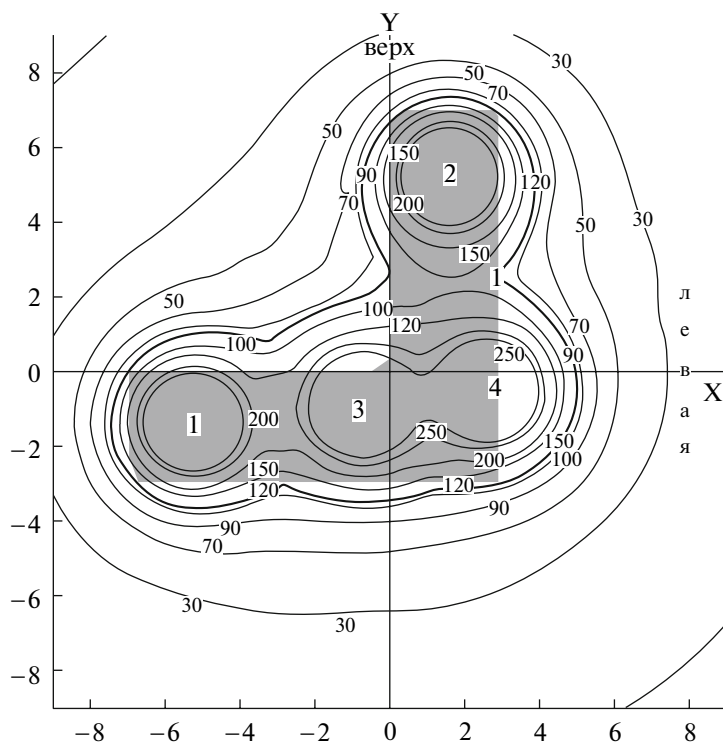


Рис. 4. Оптимальное распределение источников в невыпуклой области, которую можно назвать “бумеранг” с одинаковыми лопастями. Оптимальное размещение четырех источников не симметрично.

Полагаем номер итерации $s := 1$, задаем число источников n и начальное размещение $X_1 \subset G$.

Задаем начальный шаг смещения источников l , коэффициент его уменьшения $t \in (0, 1)$ и минимальное значение шага l_m , а также точность решения задачи ε .

В качестве начального значения глобального минимума поля действия принимаем $\mu := -\infty$.

Шаг 1. Рассчитываем поле действия источников в узлах сетки $D(X_s)$. Определяем (методом полного перебора дискретной сетки) множество A локальных минимумов поля $D(X_s)$. Вычисляем значение глобального минимума $m(X_s)$ и находим множество B глобальных (с точностью до ε) минимумов $B := \{y \in A \mid D(X_s, y) - m(X_s) \leq \varepsilon\}$.

Проверяем, выполнено ли достаточное (первое) условие окончания итерационного процесса: $A = B$ (все точки минимума являются глобальными). Если условие выполняется, то переходим к шагу 6.

Примечание. Алгоритм предполагает справедливость гипотезы о том, что равенство уровней всех локальных минимумов является достаточным для оптимальности размещения.

Шаг 2. Если $m(X_s) \geq \mu$, то запоминаем достигнутый рекорд: полагаем $\mu := m(X_s)$, $X^* := X$ и переходим к шагу 4.

Шаг 3. Уменьшаем шаг смещения: $l := lt$; если $l < l_m$ (второе условие окончания процесса), то переходим к шагу 6.

Шаг 4. Смещаем источники к точкам глобальных минимумов поля действия. Перебирая точки глобального минимума $y \in B$, находим для каждой ближайшие источники. Обозначим множество этих источников через X_y , а их расстояние до точки y — через $l(y)$. Каждую точку $x \in X_y$ смещаем в направлении $\bar{x} \bar{y}$ на расстояние $l(y)l / |x - y|$.

Шаг 5. Полагаем $s := s + 1$. Полученную на шаге 4 конфигурацию принимаем в качестве нового размещения X_s и переходим к шагу 1.

Шаг 6. Завершение расчета, выдача результатов $s, \mu, X^*, D(X^*)$.

2. Некоторые примеры. Приведем несколько примеров оптимального размещения, полученных с помощью описанного алгоритма.

На рис. 3 приводится оптимальное размещение шести источников в эллипсе с отношением осей 5 : 7. За 100% принята изолиния поля действия, отвечающая глобальному минимуму (выделена полужирной линией). Точек глобального минимума семь, из них шесть находятся на границе эллипса, и одна — в центре. При планировании внутритканевой лучевой терапии терапевтическая доза определяется по этой изолинии.

Описанный алгоритм может быть применен и к невыпуклой области. В качестве такой области была выбрана фигура, которую можно назвать “бумеранг” с одинаковыми лопастями. Для трех источников оптимальное размещение тривиально — один источник в центре и два по краям лопастей. На рис. 4 приведено оптимальное размещение четырех источников; примечательно, что оно не симметрично, как можно было бы ожидать.

Алгоритм, изложенный в п. 4.1, реализован в программном комплексе КОНТАКТ, предназначенном для планирования контактных методов лучевой терапии. Он создан автором статьи совместно с вед. научн. сотр. ЦЭМИ РАН В.Л. Ушковой.

Автор выражает благодарность В.З. Бельенкому за ряд полезных замечаний.

СПИСОК ЛИТЕРАТУРЫ

- Браиловский Н.О. (1979): Принцип максимума энтропии в теории транспортных потоков. Моделирование городских систем. Труды 1 школы-семинара. ВНИИ системных исследований.
- Фаерман Е.Ю. (1971): Проблемы долгосрочного планирования. М.: Наука.
- Фаерман Е.Ю., Клеппер Л.Я., Вейцман М.Е. (1979): Методы гравитационного и электростатического моделирования транспортных потоков и затрат применительно к задаче размещения. Моделирование городских систем. Труды 1 школы-семинара. ВНИИ системных исследований.
- Swan G.W. (1981): Lecture Notes in Biomathematics. Optimization of Human Cancer Radiotherapy. Berlin, Heidelberg, N.Y.: Springer-Verlag.

Поступила в редакцию
14.05.2008 г.