ОБ r-АЛГОРИТМАХ ШОРА

Стецюк П.И.
stetsyukp@gmail.com

Институт кибернетики им. В.М. Глушкова, Киев

VIII Московская международная конференция
по Исследованию Операций (ORM2016),
Москва, 17-22 октября 2016

Содержание

Содержание

Шор Наум Зуселевич (1937 – 2006)

Родился 1 января 1937 года в Киеве. Окончил Киевский университет имени Тараса Шевченко в 1958 году. Работал в Институте кибернетики НАН Украины с 1958 по 2006 гг. Академик НАН Украины, профессор. Автор 10 монографий и более 200 статей. Лауреат Государственных премий УССР (1973), СССР(1981), Украины (1993, 2000), премии им. В.М. Глушкова НАН Украины (1987), премии им. В.С. Михалевича НАН Украины (1997).

Статья к юбилею Н.З. Шора (январь, 2012)

75-летию со дня рождения Н.З. Шора посвящена статья

Сергиенко И.В., Стецюк П.И.

О трех научных идеях Н.З. Шора // Кибернетика и системный анализ. – 2012, № 1. – С. 4–22.

В статье описаны три центральные идеи Н.З. Шора:

обобщенный градиентный спуск (1962), использование линейных неортогональных преобразований пространства для улучшения обусловленности овражных функций (1969), двойственный подход к получению и уточнению оценок целевой функции в невыпуклых квадратичных моделях (1985).
Приведены методы и алгоритмы, разработанные на их основе в Институте кибернетики им. В.М. Глушкова НАН Украины.

Содержание

Б.Т. Поляк „Введение в оптимизацию“ (1983)

О первой идее Шора

Борис Теодорович говорит:

на стр. 128

„Основные алгоритмы минимизации гладких функций – градиентный и Ньютона – были построены на использовании линейной и квадратичной аппроксимации функции, задаваемой первыми членами ряда Тейлора. Однако, для недифференцируемой функции эта идея неприменима – такая функция не может быть хорошо аппроксимирована ни линейной, ни квадратичной функциями.“

Б.Т. Поляк „Введение в оптимизацию“ (1983)

на стр. 129

„Поэтому разработка методов минимизации негладких функций требует привлечения новых идей. Одна из них, принадлежащая Н.З. Шору, выглядит несколько неожиданно. Пишется прямой аналог градиентного метода с заменой градиента на произвольный субградиент $g_f(x)$ функции $f(x)$:

$$x_{k+1} = x_k - \gamma_k g_f(x_k). \tag{3}$$

... Значения функции в методе (3) не могут убывать монотонно. Оказывается, однако, что при этом монотонно убывает другая функция – расстояние до точки минимума, и в этом то заключается основная идея субградиентного метода (3).“

Проблемы в субградиентных методах

1. Овражная кусочно-линейная функция ($t > 1$)

$$f_1(x_1, x_2) = |x_1| + t|x_2|, \quad x^\star = (0,0), \quad f^\star = 0.$$

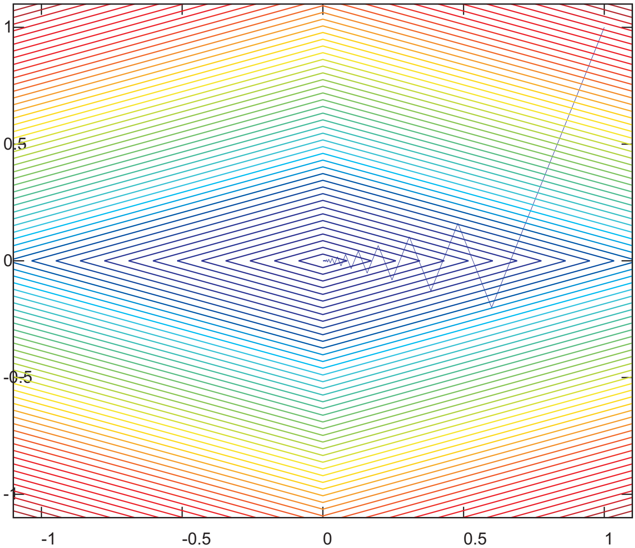
Метод Поляка (в методе (3) выбирается $\gamma_k = \frac{f(x_k)-f^*}{\|g_f(x_k)\|^2}$) сходится со скоростью геометрической прогрессии со знаменателем $q(t) = \sqrt{1 - \frac{1}{t^2}}$ (Поляк,1969).

2. Существенно овражная кусочно-квадратичная функция

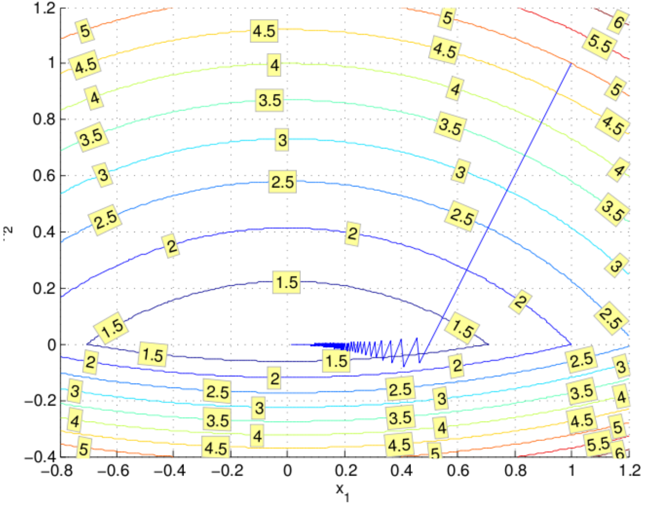
$$f_2(x_1, x_2) = \max\left\{x_1^2 + (2x_2 - 2)^2 - 3, x_1^2 + (x_2 + 1)^2\right\},$$

Вырождение в точке минимума $x^\star = (0,0)$, $f^\star = 1$.

Метод Поляка для $f_1(x_1, x_2) = |x_1| + 10|x_2|$



Метод Поляка для $f_2(x_1, x_2)$ (10000 итераций)



Содержание

Вторая идея Шора (1969)

состоит в

применении линейных неортогональных преобразований пространства для улучшения свойств минимизируемой функции в преобразованном пространстве переменных.

Субградиентный метод Shor69

Пусть x_0 – начальное приближение, B_0 – $n \times n$-матрица. Итерация субградиентного метода с последовательным преобразованием пространства переменных имеет вид:

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}, \quad B_{k+1} = B_k T_k, \qquad \text{(Shor69)}$$

где h_k – шаговый множитель, T_k – $n \times n$-матрица.

Если в методе (Shor69) матрицы T_k подбирать так, чтобы в преобразованном пространстве поверхности овражных функций становились менее овражными, то такой метод окажется эффективнее, чем субградиентный метод.

Об двух семействах методов с растяжением

Это подтверждают два семейства методов Н.З. Шора:

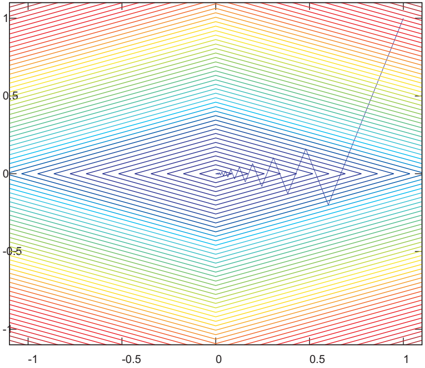
- субградиентные методы с растяжением пространства в направлении субградиента. Их частный случай – метод эллипсоидов (Юдин, Немировский, 1976), (Шор, 1977);
- субградиентные методы с растяжением пространства в направлении разности двух последовательных субградиентов (r-алгоритмы).

Методы используют оператор растяжения пространства

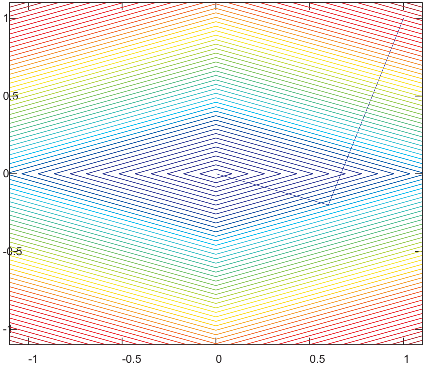
$$R_\alpha(\xi) = I_n + (\alpha - 1)\xi\xi^T, \qquad \text{где} \qquad \alpha > 1, \qquad (*)$$

I_n – единичная $n{\times}n$-матрица, вектор $\xi \in E^n$ такой, что $\|\xi\|{=}1$.

Ускорение по Шору для $f_1(x_1, x_2)=|x_1|+10|x_2|$

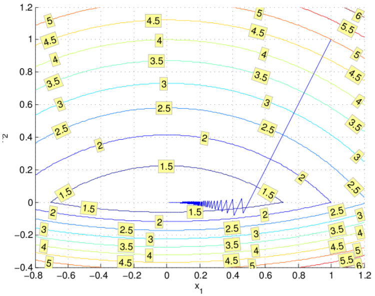


Метод **Поляка**

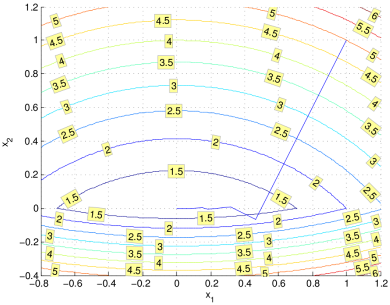


Метод **amsg2** (2 итерации)

Ускорение по Шору для $f_2(x_1, x_2)$



Метод **Поляка** (10000 итер.)



Метод **amsg2** (31 итерация)

Содержание

История r-алгоритмов

r-алгоритмы получили свое название от русского слова "разность". Их называют r-алгоритмами Шора, либо r-алгоритмами Шора–Журбенко.

Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С.51–59.

Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. – Киев: Наукова думка, 1979.

В r-алгоритмах шаг выбирается из условия минимума функции по направлению сдвига, благодаря чему определяются те два последовательных субградиента, растяжение по „разности“ которых улучшает свойства овражной функции в преобразованном пространстве переменных.

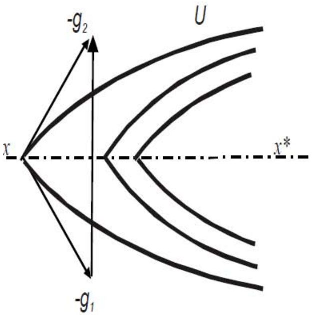
Содержание

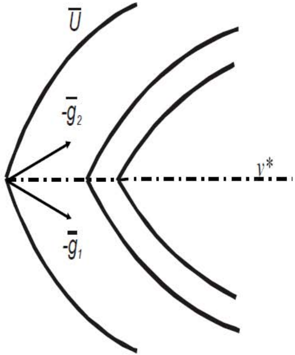
Идея r-алгоритмов

В r-алгоритмах шаг выбирается из условия точного (приближенного) минимума функции по направлению сдвига, благодаря чему определяются те два последовательных субградиента, растяжение по разности которых улучшает свойства овражной функции в преобразованном пространстве переменных.

При определенной регулировке шага и коэффициентов растяжения пространства r-алгоритмы являются монотонными (или почти монотонными) по минимизируемой функции.

Идея r-алгоритмов (по рисунку из книг Шора)





или почему возможно построить алгоритмы,

монотонные по значениям минимизируемой функции.

Вычислительная схема r-алгоритмов

r-Алгоритмом называется процедура построения последовательностей $\{x_k\}_{k=0}^{\infty}$ и $\{B_k\}_{k=0}^{\infty}$ по правилу:

$$x_{k+1} = x_k - h_k B_k \xi_k, \ \ B_{k+1} = B_k R_{\beta_k}(\eta_k), \ \ k = 0, 1, 2, \dots, \qquad (1)$$

$$\text{где} \quad \xi_k = \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}, \quad h_k \geq h_k^{\star} = \operatorname*{argmin}_{h \geq 0} f(x_k - hB_k\xi_k), \qquad (2)$$

$$\eta_k = \frac{B_k^T r_k}{\|B_k^T r_k\|}, \ \ r_k = g_f(x_{k+1}) - g_f(x_k), \quad \beta_k = \frac{1}{\alpha_k} < 1. \qquad (3)$$

Здесь x_0 — начальное приближение, $B_0 = I_n$ – единичная $n \times n$-матрица, h_k – шаговый множитель, α_k – коэффициент растяжения пространства, $g_f(x_k)$ и $g_f(x_{k+1})$ – произвольные субградиенты функции $f(x)$ в точках x_k и x_{k+1}.

Содержание

Два результата для гладких функций

Для минимизации гладких функций r-алгоритмы по своей структуре близки к алгоритмам квазиньютоновского типа с переменной метрикой. Это определило такие результаты.

1. Предельный вариант r-алгоритма с бесконечным коэффициентом растяжения (здесь $\beta = 0,\ h_k = h_k^*$) является проективным вариантом метода сопряженных градиентов (Шор, Журбенко, 1971).

2. Предельный вариант r-алгоритма с восстановлением матрицы B_k после каждых n итераций обладает квадратичной скоростью сходимости при обычных условиях гладкости и регулярности $f(x)$ (там же, 1971).

Самый общий результат

Теорема (Шор, 1975)

для класса почти дифференцируемых кусочно-гладких функций определяет достаточные условия, при которых $r_\mu(\alpha)$-алгоритм $(h_k{=}h_k^*)$ сходится к локальному минимуму.

Эти условия слишком сильные и не выполняются даже для кусочно-линейных функций. Наиболее типичная ситуация, когда это происходит, связана с нарушением линейной независимости множества $G_f(x)$ – множества почти-градиентов $f(x)$ в точке x.

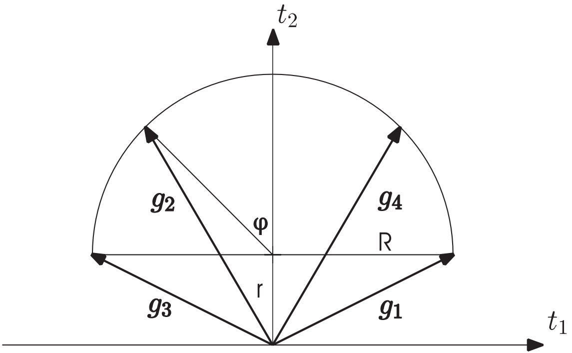
Можно ли улучшить результат Шора (1975)?

Оказывается нельзя.

Построен пример выпуклой кусочно-линейной функции, для которой возможно "зацикливание" $r_\mu(\alpha)$-алгоритма.

Стецюк П.И. К вопросу сходимости r-алгоритмов // Кибернетика и систем. анализ. – 1995. – № 6. – С. 173–177.

Идея „ловушек“ для $r_\mu(\alpha)$-алгоритма



Векторы $r_1 = g_2 - g_1$, $r_2 = g_3 - g_2$, $r_3 = g_4 - g_3$, $r_4 = g_1 - g_4$ такие, что r_1 и r_2, r_3 и r_4 – ортогональны.

После четырех последовательных растяжений пространства с коэффициентом α картина векторов останется подобной, а длины векторов уменьшатся в α^2 раз.

Три статьи Шора о сходимости r-алгоритмов

- Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С. 51–59.

- Шор Н.З. Исследование сходимости метода градиентного типа с растяжением пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1975. – № 4. – С. 48-53.

- Шор Н.З. Монотонные модификации r-алгоритмов и их приложения // Кибернетика и системный анализ. – 2002. – № 6. – С. 74–96.

Содержание

$r(\alpha)$-алгоритм с адаптивным шагом

Оказался эффективным вариантом r-алгоритмов и на его основе разработано ряд программных реализаций (Журбенко Н.Г., Кунцевич А.В., Лиховид А.П. и др.)

Здесь величина шага h_k адаптивно настраивается с помощью параметров h_0, q_1, n_h, q_2, где h_0 – величина начального шага (используется на 1-й итерации, на каждой последующей итерации уточняется); q_1 – коэффициент уменьшения шага $(q_1 \leq 1)$, если условие завершения спуска по направлению $(h_k > h_k^*)$ выполняется всего за один шаг одномерного спуска; q_2 – коэффициент увеличения шага $(q_2 \geq 1)$; натуральное число n_h задает число шагов одномерного спуска $(n_h > 1)$, через каждые из которых шаг в одномерном спуске будет увеличиваться в q_2 раз. **(Шор, Стеценко, 1989)**

Octave-функция ralgb5 (Стецюк, 2011)

```
%     ralgb5 реализует $r(\alpha)$-алгоритм с адаптивным шагом,
%     использует подготовленную пользователем octave-функцию
%     function [f,g] = calcfg(x), которая вычисляет значения
%     функции $f=f(x)$ и её субградиента $g(x)$ в точке $x$.

%  Входные параметры:
%    calcfg -- имя функции calcfg(x) для вычисления f и g
%    x  -- начальная точка x0(1:n) (на выходе портится)
%    alpha -- коэффициент растяжения пространства
%    h0, nh, q1, q2 -- параметры адаптивной регулировки шага
%    epsx, epsg, maxitn -- параметры останова

%  Выходные параметры:
%    xr -- найденная точка минимума xr(n)
%    fr -- значение функции в точке минимума
%    itn -- число затраченных итераций
%    ncalls -- число вызовов функции calcfg
%    istop -- код останова (2=epsg,3=epsx,4=maxitn,5=error)
```

Код octave-функции ralgb5 (Стецюк, 2011)

```
function [xr,fr,itn,ncalls,istop]=ralgb5(calcfg,x,alpha,h0,q1,
                                          q2,nh,epsg,epsx,maxitn);
itn=0; hs=h0; B=eye(length(x)); xr=x;                        # row001
ncalls = 1; [fr,g0] = calcfg(xr);                            # row002
printf("itn %4d f %14.6e fr %14.6e ls %2d ncalls %4d\n",     # row003
       itn, fr, fr, 0, ncalls);
if(norm(g0) < epsg) istop = 2;  return; endif                # row004
for (itn = 1:maxitn)                                         # row005
   dx = B * (g1 = B' * g0)/norm(g1);                         # row006
   d = 1; ls = 0; ddx = 0;                                   # row007
   while (d > 0)                                             # row008
      x -= hs * dx; ddx += hs * norm(dx);                    # row009
      ncalls ++; [f, g1] = calcfg(x);                        # row010
      if (f < fr)  fr = f; xr = x; endif                     # row011
      if(norm(g1) < epsg) istop = 2;  return; endif          # row012
      ls ++; (mod(ls,nh)==0) && (hs *= q2);                  # row013
      if(ls > 500) istop = 5;  return; endif                 # row014
      d = dx' * g1;                                          # row015
   endwhile                                                  # row016
   (ls == 1) && (hs *= q1);                                  # row017
   printf("itn %4d f %14.6e fr %14.6e ls %2d ncalls %4d\n",  # row018
          itn, f, fr, ls, ncalls);
   if(ddx < epsx) istop = 3;  return; endif                  # row019
   xi = (dg = B' * (g1 - g0) )/norm(dg);                     # row020
   B += (1 / alpha - 1) * B * xi * xi';                      # row021
   g0 = g1;                                                  # row022
endfor                                                       # row023
istop = 4;                                                   # row024
endfunction
```

Выбор параметров для ralgb5

При минимизации негладких функций рекомендуется:

$$\alpha = 2 \div 3,\ h_0 = 1.0,\ q_1 = 1.0,\ q_2 = 1.1 \div 1.2,\ n_h = 2 \div 3.$$

Если известна априорная оценка расстояния от начальной точки x_0 до точки минимума x^*, то начальный шаг h_0 целесообразно выбирать порядка $\|x_0 - x^*\|$.
При минимизации гладких функций рекомендуется

$$q_1 = 0.8 \div 0.95.$$

При таком выборе параметров, как правило, число спусков по направлению редко превосходит два, а за n шагов точность по функции улучшается в три-пять раз.

Содержание

Актуально и сегодня

Н.З.Шор, В.И.Гершович, 1982

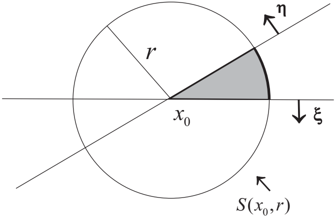
„Теория всего класса алгоритмов с растяжением пространства далека от совершенства. Нам кажется достаточно реалистичной целью – построение такого алгоритма, который по своей практической эффективности не уступал бы r-алгоритму и был столь же хорошо обоснован как метод эллипсоидов“.

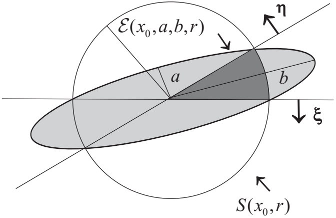
Одна из попыток реализована в книге

Стецюк П.И. Методы эллипсоидов и r-алгоритмы. – Кишинэу: Эврика, 2014. – 488 с.

где используется антиовражный прием, близкий к тому, который имеет место в r-алгоритмах.

Тело W и Специальный эллипсоид





Тело W получено как пересечение шара и двух полупространств.

Специальный эллипсоид содержит W и имеет минимальный объем.

Стецюк П.И. r-алгоритмы и эллипсоиды // Кибернетика и системный анализ. – 1996. – № 1. – С. 113–134.

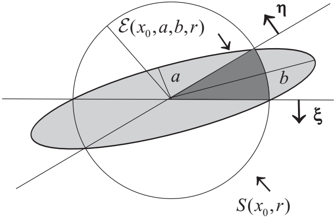
Свойства Специального эллипсоида

1. Если угол φ между векторами ξ и η тупой, то эллипсоид содержит тело W. Объем эллипсоида меньше, чем объем шара, и это уменьшение равно

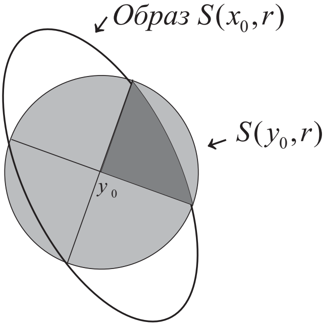
$$\sqrt{1-(\xi,\eta)^2}=\sqrt{1-(\cos\varphi)^2}.$$

2. Преобразовать эллипсоид в шар можно растяжением пространства в направлении $\frac{\xi-\eta}{\|\xi-\eta\|}$ с коэффициентом $\alpha_1=\frac{1}{\sqrt{1+(\xi,\eta)}}$ и сжатием пространства в ортогональном направлении $\frac{\xi+\eta}{\|\xi+\eta\|}$ с коэффициентом $\alpha_2=\frac{1}{\sqrt{1-(\xi,\eta)}}$.

Эллипсоид до и после растяжения



Специальный эллипсоид



в дважды растянутом пространстве становится шаром

Близость к r-алгоритмам

В преобразованном пространстве эллипсоид станет шаром, а образы векторов ξ и η будут ортогональными.

Это позволяет “расширить” конус подходящих направлений убывания функции для субградиентного процесса в преобразованном пространстве переменных, аналогично тому, как это делается в r-алгоритмах.

Растяжение пространства реализуется в направлении разности двух нормированных субградиентов и близким к направлению разности двух субградиентов оно будет только тогда, когда нормы субградиентов близки.

Заключение

Актуально и сегодня

„Теория всего класса алгоритмов с растяжением пространства далека от совершенства. Нам кажется достаточно реалистичной целью – построение такого алгоритма, который по своей практической эффективности не уступал бы r-алгоритму и был столь же хорошо обоснован как метод эллипсоидов“.

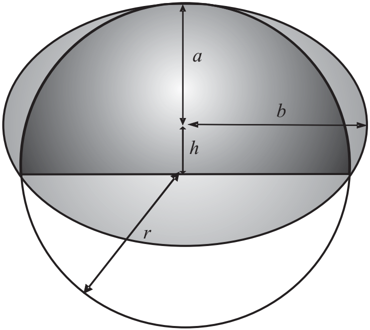
Thanks

Volkswagen Foundation

Вопросы?

СПАСИБО ЗА ВНИМАНИЕ!

BACK UP SLIDES: Оптимальный 1d-эллипсоид



Эллипсоид $\mathcal{E}_n$, содержащий полушар в E^n, имеет минимальный обьем, если

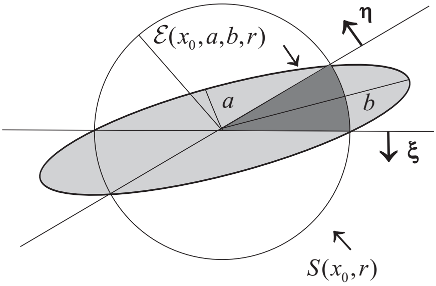
$$a = \frac{n}{n+1} r, \ b = \frac{n}{\sqrt{n^2-1}} r, \ h = \frac{1}{n+1} r.$$

Чтобы преобразовать $\mathcal{E}_n$ в шар нужно растянуть пространство с коэффициентом $\alpha = \frac{b}{a} = \sqrt{\frac{n+1}{n-1}}$.

На каждой итерации МЭ объем эллипсоида уменьшается в

$$q(n) = \frac{\mathrm{vol}(\mathcal{E}_n)}{\mathrm{vol}(S_n)} = \frac{a}{r}\left(\frac{b}{r}\right)^{n-1} = \sqrt{\frac{n-1}{n+1}}\left(\frac{n}{\sqrt{n^2-1}}\right)^n \approx 1-\frac{1}{2n},$$

BACK UP SLIDES: Оптимальный 2d-эллипсоид



Преобразование в шар требует растяжения

в направлении $\frac{\xi-\eta}{\|\xi-\eta\|}$

с коэф. $\alpha_1 = \frac{1}{\sqrt{1+(\xi,\eta)}} > 1$

и последующего сжатия

в направлении $\frac{\xi+\eta}{\|\xi+\eta\|}$

с коэф. $\alpha_2 = \frac{1}{\sqrt{1-(\xi,\eta)}} < 1.$

$$q = \frac{\mathrm{vol}(\mathcal{E}(x_0,a,b,r))}{\mathrm{vol}(S(x_0,r))} = \left(\frac{a}{r}\right)\left(\frac{b}{r}\right) = \sqrt{1-(\xi,\eta)^2}.$$