МЕТОД АППРОКСИМАЦИИ ШЕПАРДА И ЕГО ПРИЛОЖЕНИЯ

. . .

SHEPARD APPROXIMATION METHOD AND IT'S APPLICATIONS

Горнов А.Ю. gornov@icc.ru

ИДСТУ СО РАН, Иркутск

Горнов А.Ю.

Зароднюк Т.С.

Аникин А.С.

Финкельштейн Е.А.

Доржиева А.Б.

Веялко И.А.

Никифоров А.В.

Постановка задачи

$$\{x_i, f_i\}, \quad i = \overline{1, k}$$

$$x_i \in \mathbb{R}^n$$

$$f_i \in \mathbb{R}$$

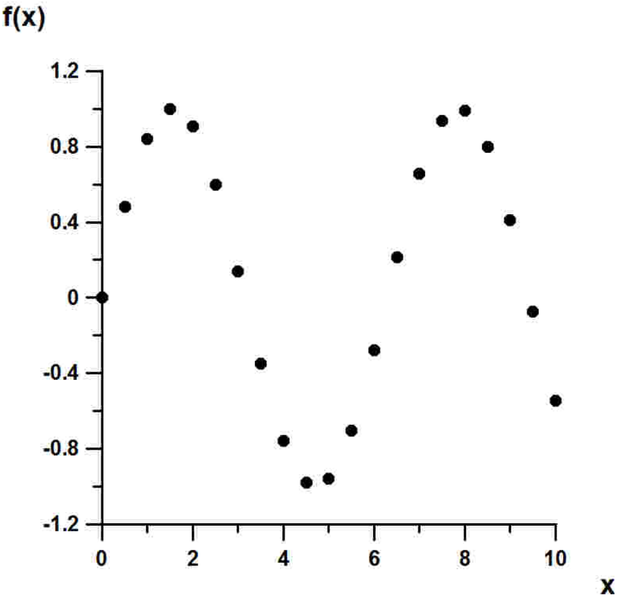
Аппроксимант Шепарда

$$F(x) = \begin{cases} f_i & \text{если } x = x_i \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x-x_i\|^4}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4}} & \text{если } x \neq x_i \end{cases}$$

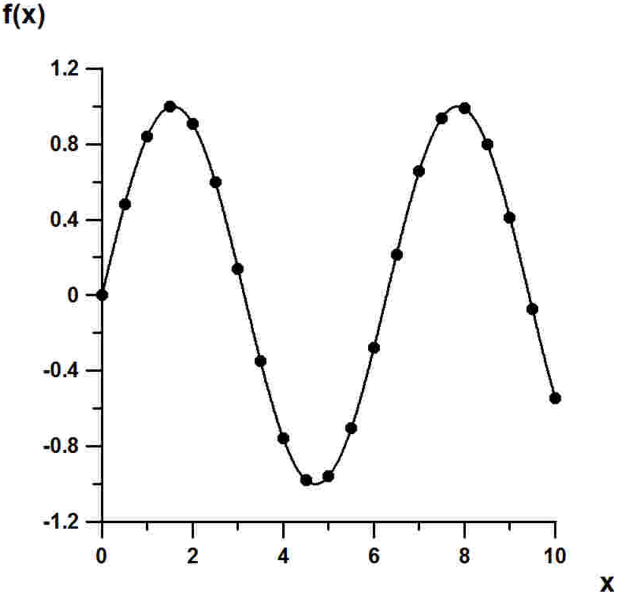
Аппроксимант Шепарда

- Без использования производных
- С использованием производных
- Методика сглаживания
- Методика овыпукления
- Методика восстановления производных

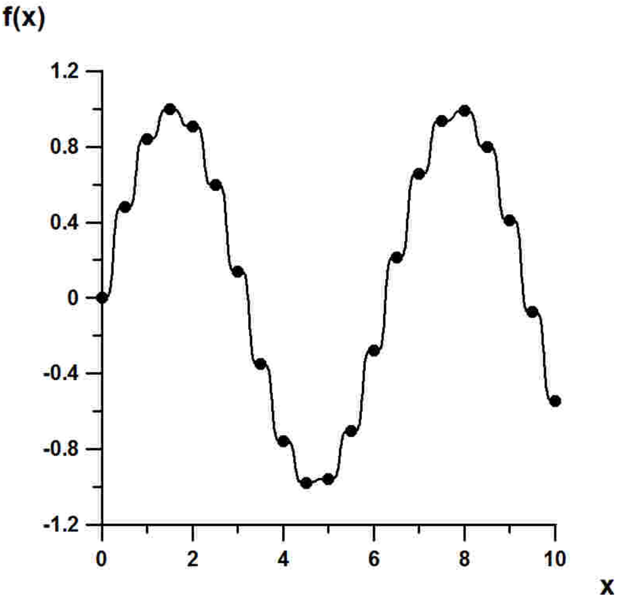
Аппроксимант Шепарда



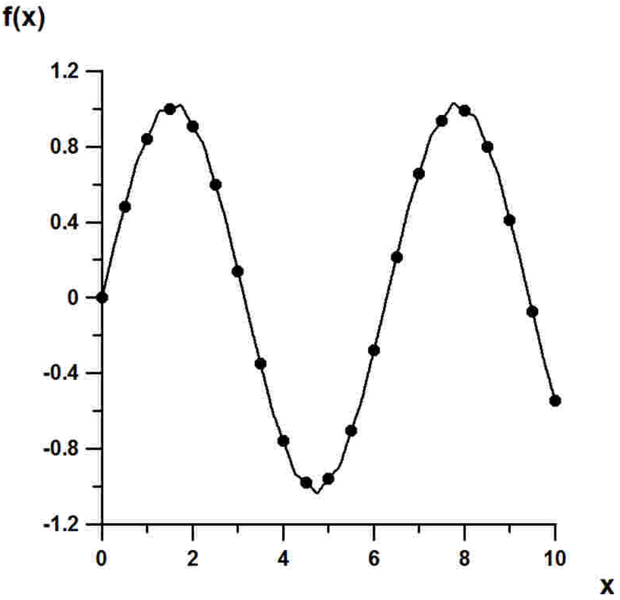
Аппроксимант Шепарда



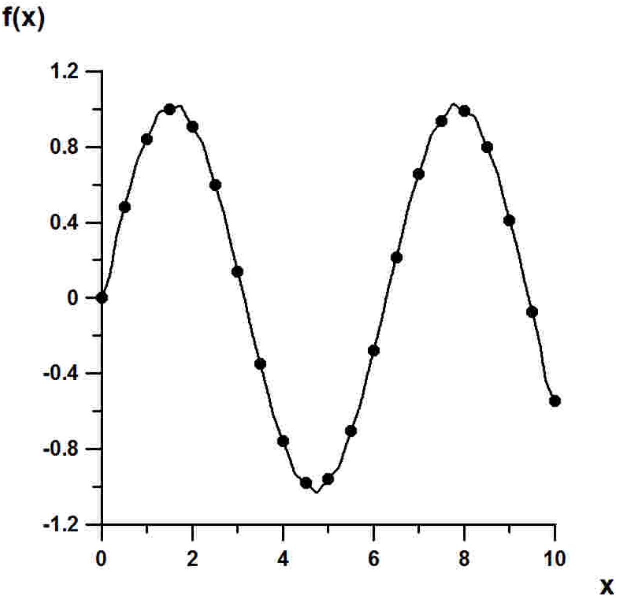
Аппроксимант Шепарда



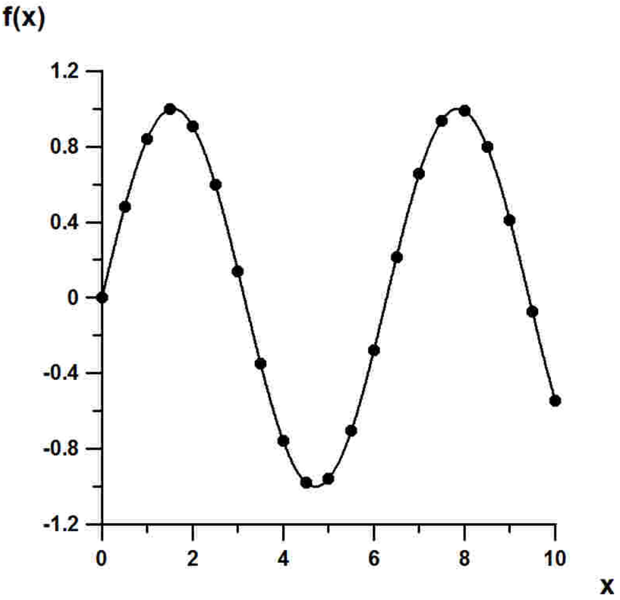
Аппроксимант Шепарда



Аппроксимант Шепарда



Аппроксимант Шепарда



Свойства аппроксиманта

- Интерполяция данных
- Гарантированное отсутствие полюсов
- Непрерывная дифференцируемость
- Произвольная размерность функции
- Неравномерная (нерегулярная) сетка
- Быстрая вычисляемость
  ($4 \cdot k \cdot n + 1$ операций)

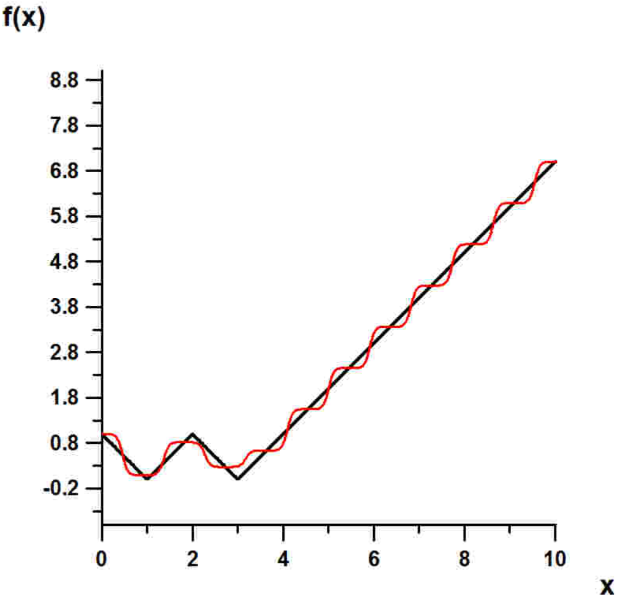
Время вычисления (сек.)

Точек / переменных	10^2	10^3	10^4	10^5	10^6
10^2	0.001	0.001	0.008	0.083	0.850
10^3	0.001	0.008	0.085	0.820	
10^4	0.008	0.082	0.820		
10^5	0.084	0.850			
10^6	0.900				

Недифферецируемая функция

$$F(x) = \Big| x - \big| x - 1 + |x-2| \big| \Big|$$

Недифферецируемая функция

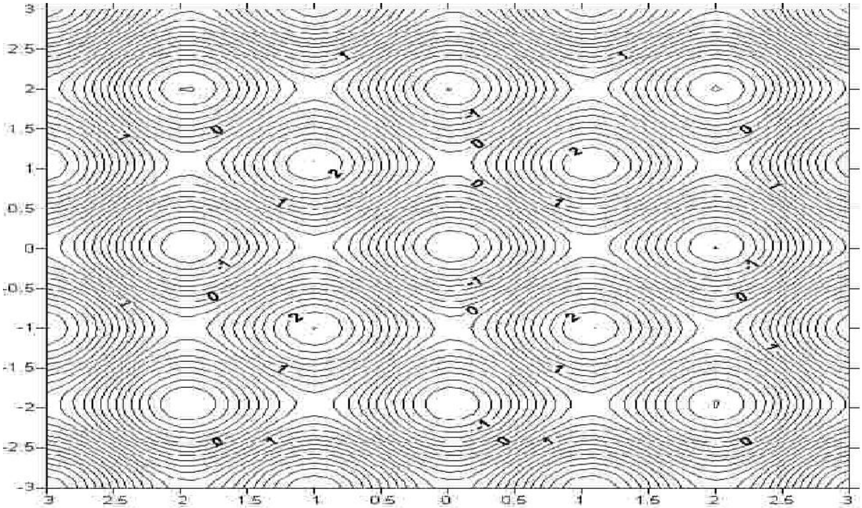


Тестовая задача

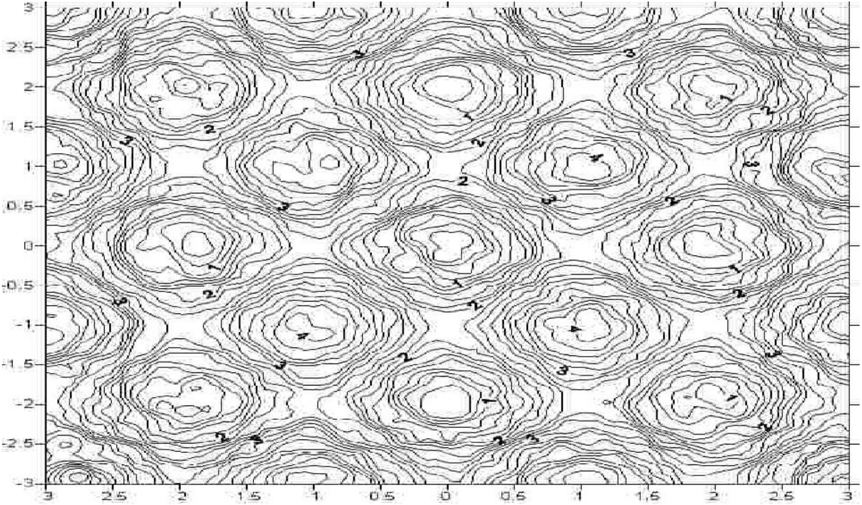
$$f(x) = \sum_i sin(\pi \cdot x_i - 0.5\pi) + 0.1 \sum_i x_i^2$$

$$\begin{aligned} -3 \leq x_i \leq 3 \\ x^* = 0 \\ f* = 0 \end{aligned}$$

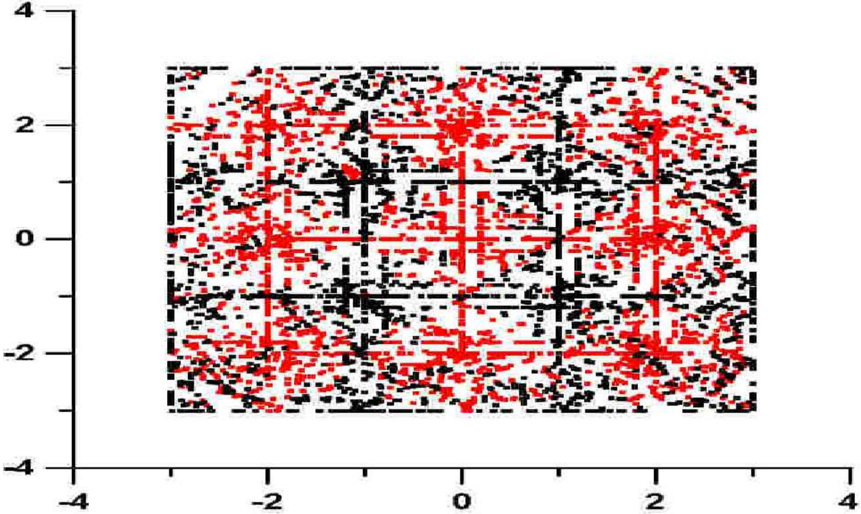
Тест. Полная картина



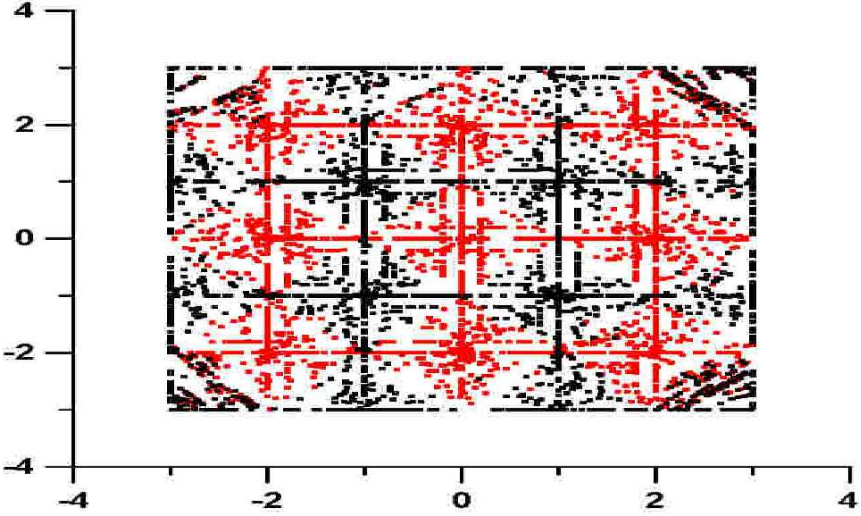
Тест. 2D Шепард-проекция



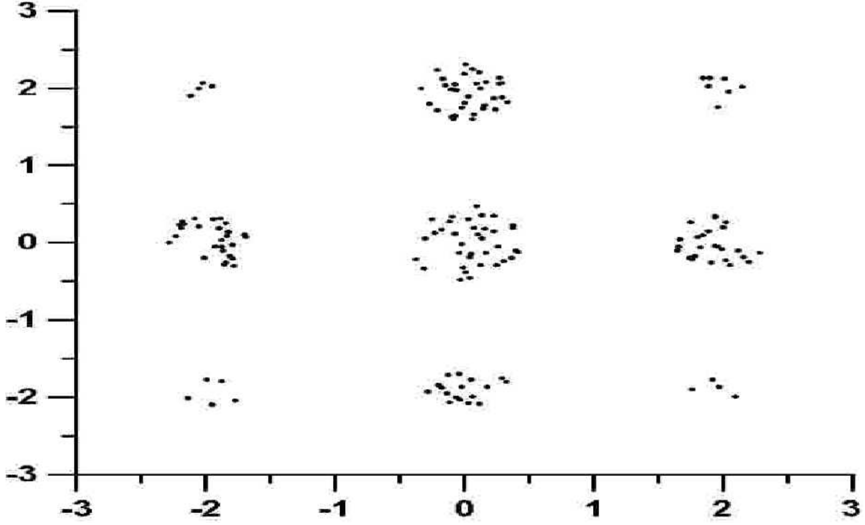
Тест. Карта хребтов. 2000 проб



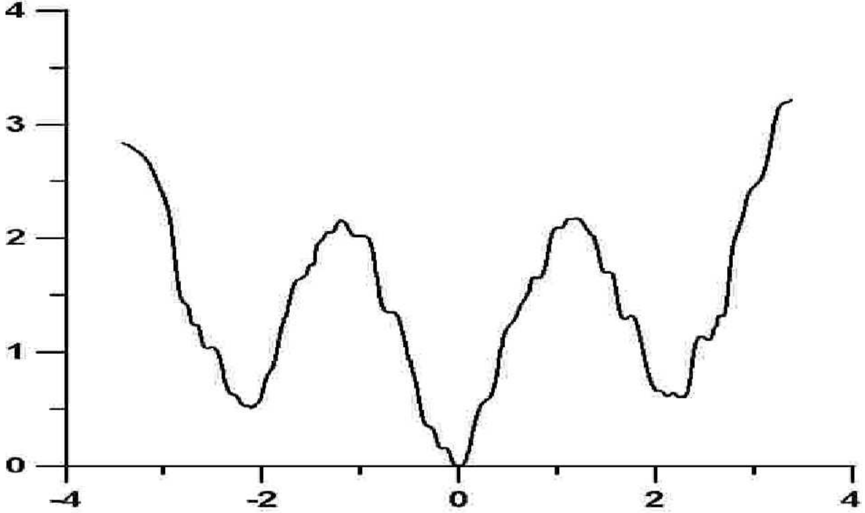
Тест. Карта хребтов. 20000 проб



Тест. Карта сгущений. 2000 проб



Тест. Линейный поиск. 2000 проб



Метод с использованием производной

$$F_p(x) = \begin{cases} f_i \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i + f_i'(x - x_i)}{\|x - x_i\|^4}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4}} \end{cases}$$

$f_i'(x)$ — вектор градиента в узле x_i

Методика сглаживания

$$F_c(x) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x-x_i\|^4+C}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4+C}} \end{cases}$$

$$C \neq 0$$

Методика сглаживания

$$\lim_{C\to\infty}\frac{\sum\limits_{i=1}^{k}\frac{f_i}{\|x-x_i\|^4+C}}{\sum\limits_{i=1}^{k}\frac{1}{\|x-x_i\|^4+C}}=\frac{\frac{\sum_{i=1}^{k}f_i}{C}}{\frac{k}{C}}=\frac{\sum\limits_{i=1}^{k}f_i}{k}=\overline{f}$$

Методика сглаживания

С целью улучшения свойств приближения в операторе Шепарда, используем информацию о производных в узловых точках:

$$F_c(x) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i + f'(x - x_i)}{\|x - x_i\|^4 + C}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C}} \end{cases}$$

Методика сглаживания

Наличие производных, в некоторых случаях, может нарушать “сглаживание”. Для гарантированного сохранения этого эффекта заменим f' на приближение z также с помощью аппроксиманта Шепарда

Методика сглаживания

$$F_c(x) = \begin{cases} f_i \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i + z_i(x - x_i)}{\|x - x_i\|^4 + C_1}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C_1}} \end{cases}$$

где $z_i = z(x_i)$ - функция приближения производной

Методика сглаживания

$$z(x) = \frac{\sum\limits_{i=1}^{k} \frac{f_i'(x)}{\|x-x_i\|^4+C_2}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4+C_2}}$$

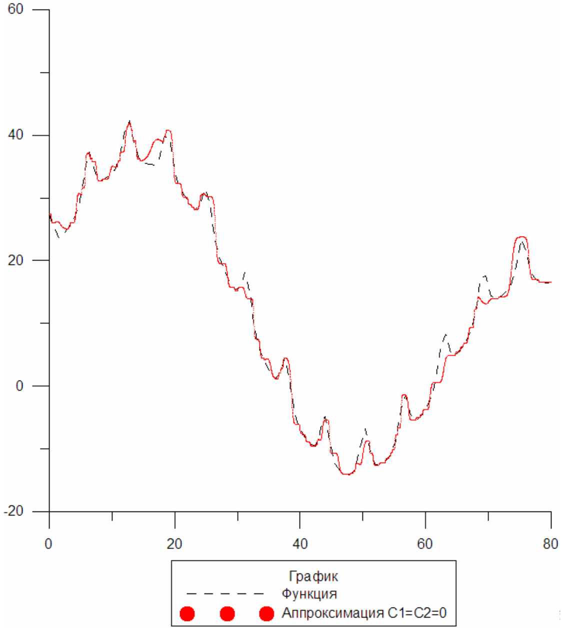
Методика сглаживания

$$f(x) = 20 + 20sin(0.1x) - 0.3x + e^{2cos(x)}$$

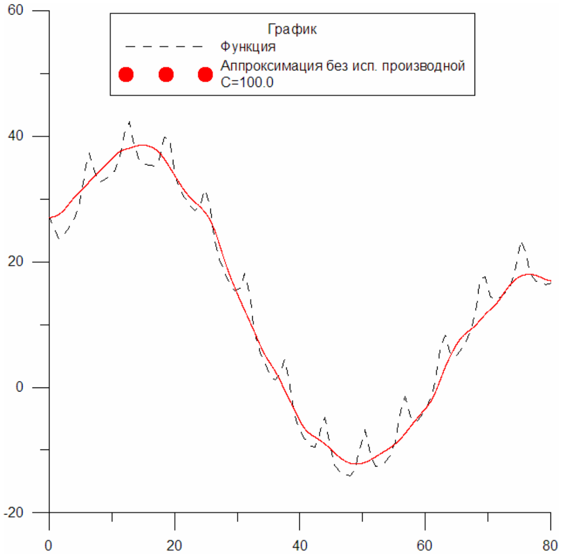
$$x \in [0, 80]$$

Аппроксимация по 100 случайным точкам

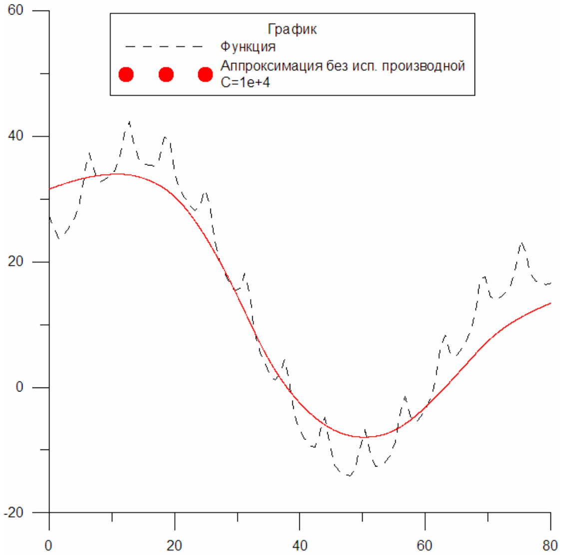
Методика сглаживания



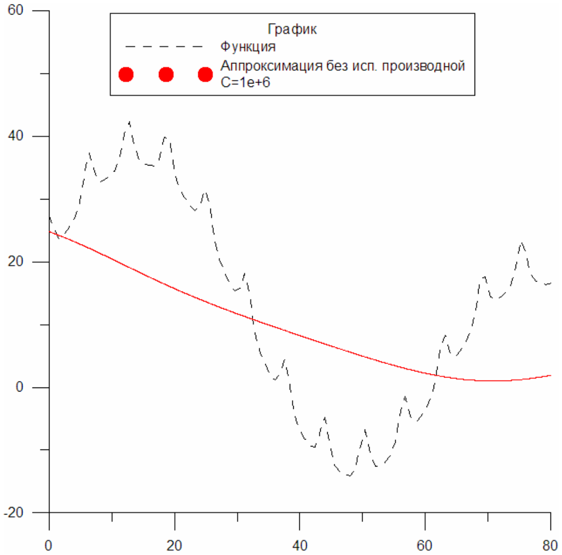
Методика сглаживания



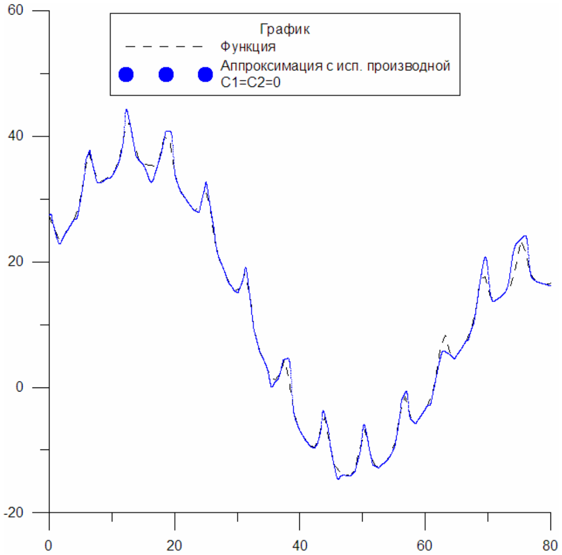
Методика сглаживания



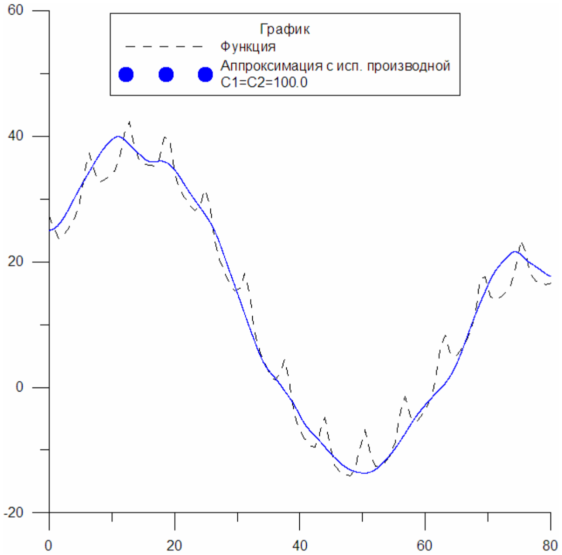
Методика сглаживания



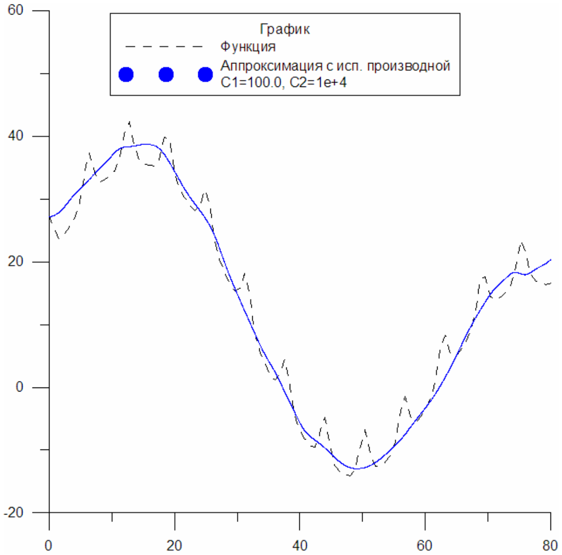
Методика сглаживания



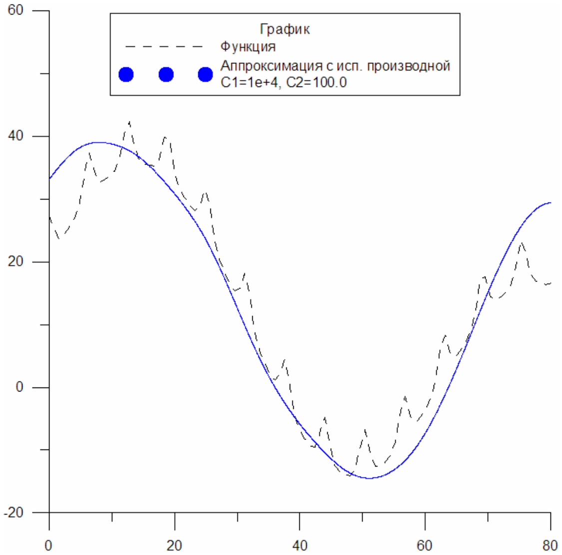
Методика сглаживания



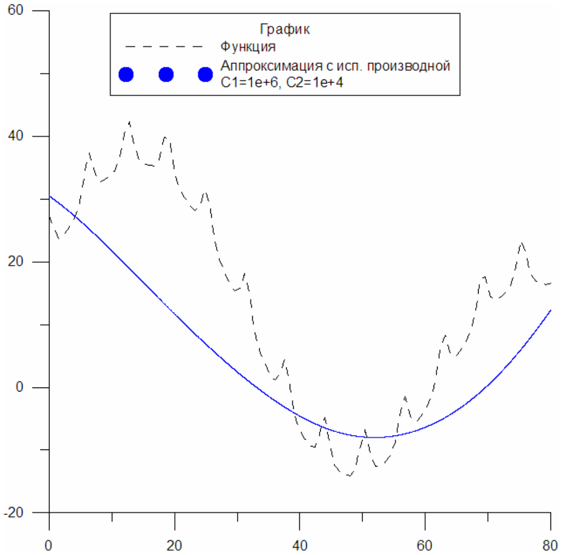
Методика сглаживания



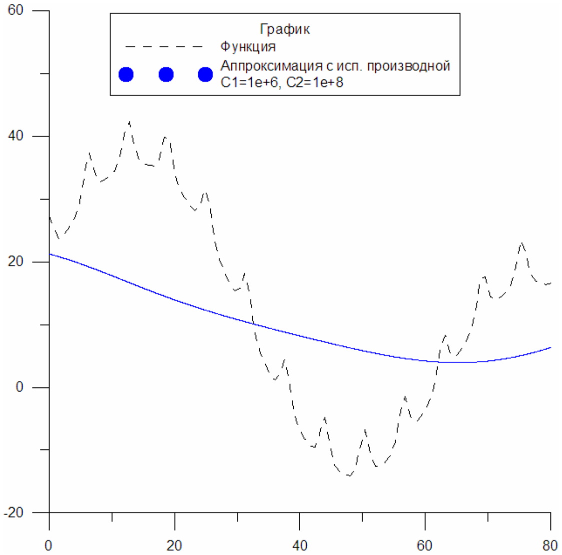
Методика сглаживания



Методика сглаживания



Методика сглаживания



Методика сглаживания

$$S(x, C) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x - x_i\|^4 + C}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C}} \end{cases}$$

$$C \neq 0$$

Методика сглаживания

$$\Phi = \int_{t_0}^{t_1} f(x)dx - \beta \int_{t_0}^{t_1} S(x, C)dx$$

$f(x)$ — сглаживаемая функция

t_0, t_1 — границы отрезка, на котором задана функция

β — параметр, обозначающий отношение площадей исходной и сглаженной функций

Методика сглаживания

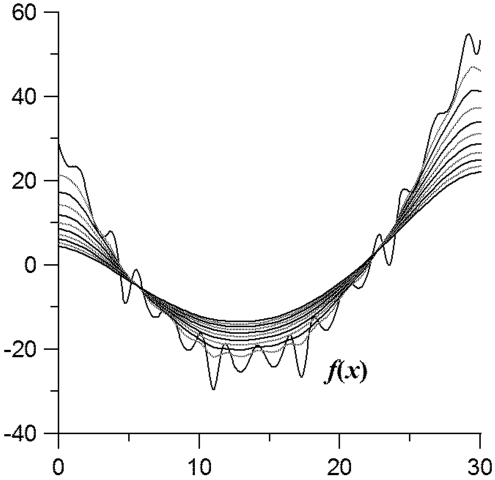
$$f(x) = 30 - 8x + 0.3x^2 + 6\sin(x + 3\cos(x)) + \varepsilon$$

$x \in [0, 30]$

$|\varepsilon| \leq 0.2$

Методика сглаживания

Зашумленная функция и сглаживающие аппроксимирующие функции в порядке возрастания β.



Методика сглаживания

На основе решения ряда тестовых задач можно утверждать, что задание параметра $\beta \in [1.1,\ 1.2]$ может дать наиболее точный результат

Методика сглаживания

- При применении методики сглаживания оператор Шепарада перестанет быть интерполянтом (становится аппроксимантом)

- При неограниченном увеличении C, $F(x)$ стремится к среднему значению табличных значений функции

Методика овыпуклення

$$F_v(x) = \begin{cases} f_i \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x-x_i\|^4 + C(f_i - f_{min})}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4 + C(f_i - f_{min})}} \end{cases}$$

f_{min} — минимальное значение функции из набора узловых значений

Методика овыпуклления

$$F_v(x) = \begin{cases} f_i \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i + f'(x - x_i)}{\|x - x_i\|^4 + C(f_i - f_{min})}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C(f_i - f_{min})}} \end{cases}$$

Методика овыпукления

$$F_v(x) = \begin{cases} f_i \\ \frac{\sum\limits_{i=1}^{k} \frac{f_i + z_i(x - x_i)}{\|x - x_i\|^4 + C_1(f_i - f_{min})}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C_1(f_i - f_{min})}} \end{cases}$$

где $z_i = z(x_i)$ - функция приближения производной

Методика овыпуклениия

$$z(x) = \frac{\sum\limits_{i=1}^{k} \frac{f_i'(x)}{\|x - x_i\|^4 + C_2}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4 + C_2}}$$

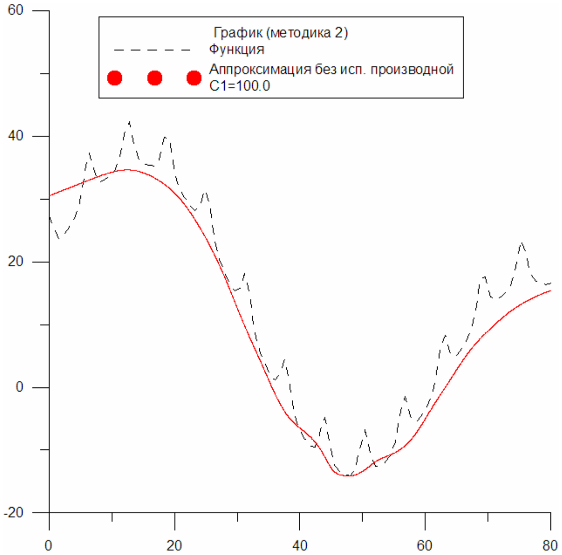
Методика овыпукления

$$f(x) = 20 + 20sin(0.1x) - 0.3x + e^{2cos(x)}$$

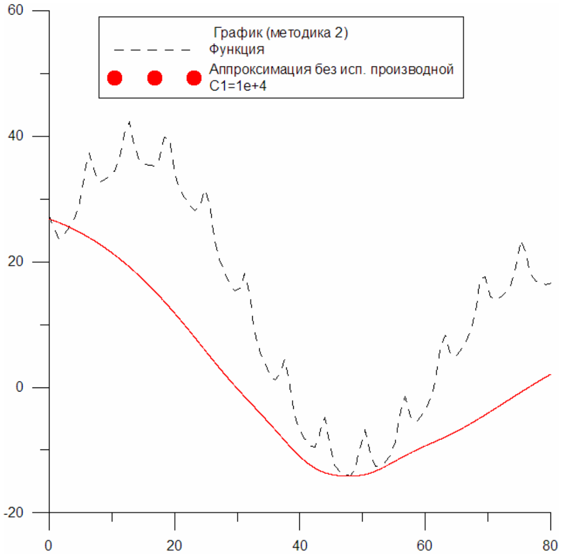
$$x \in [0, 80]$$

Аппроксимация по 100 случайным точкам

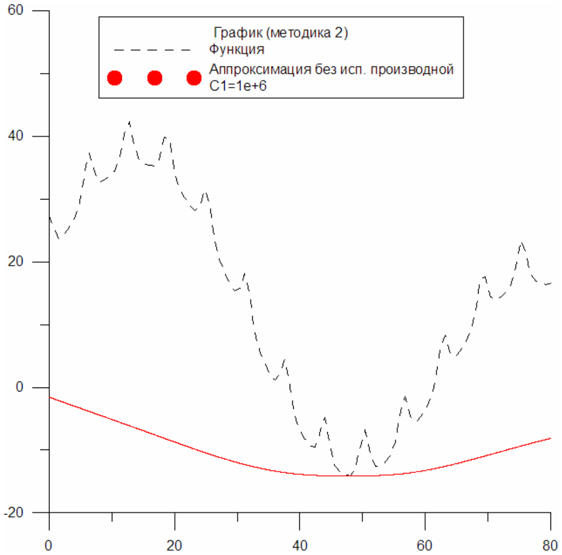
Методика овыпукления



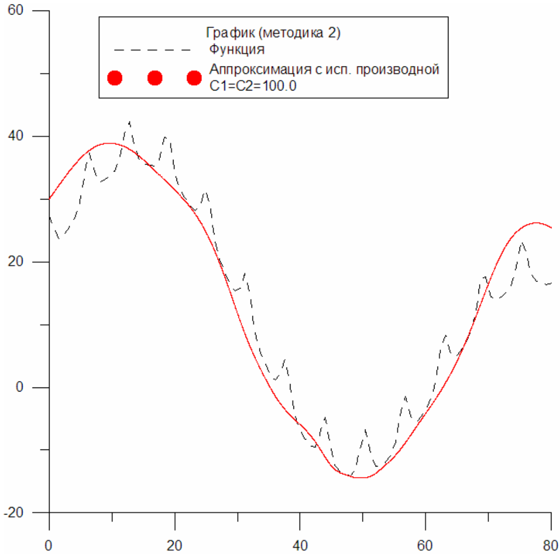
Методика овыпукления



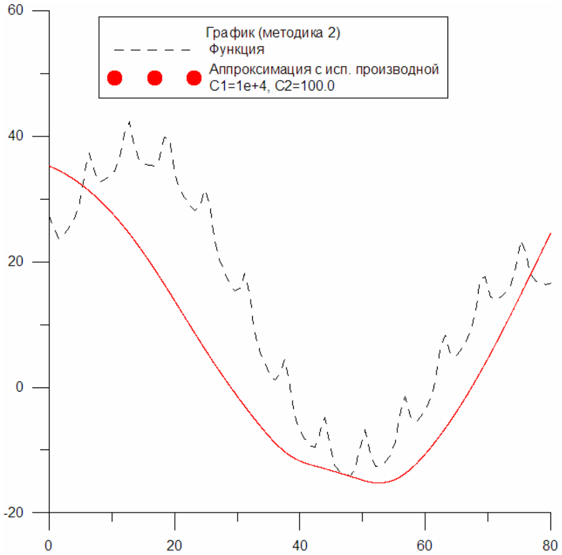
Методика овыпукления



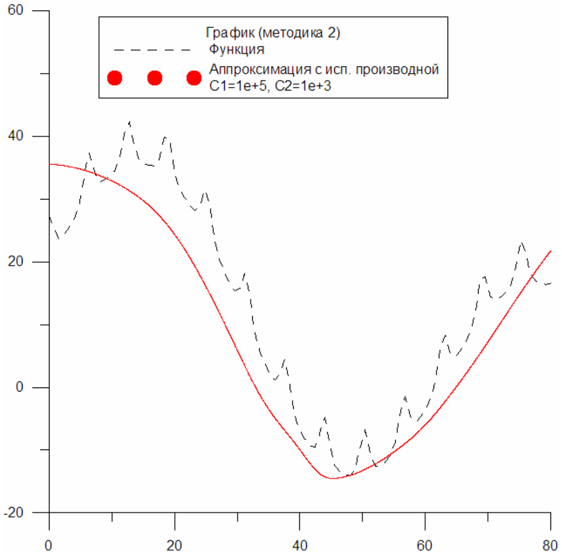
Методика овыпукления



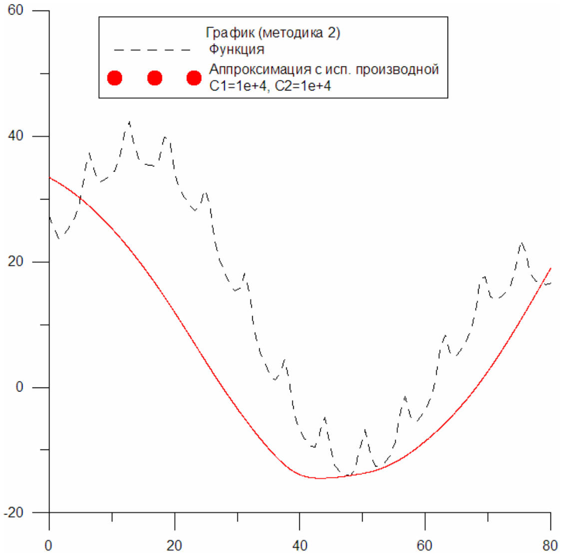
Методика овыпукления



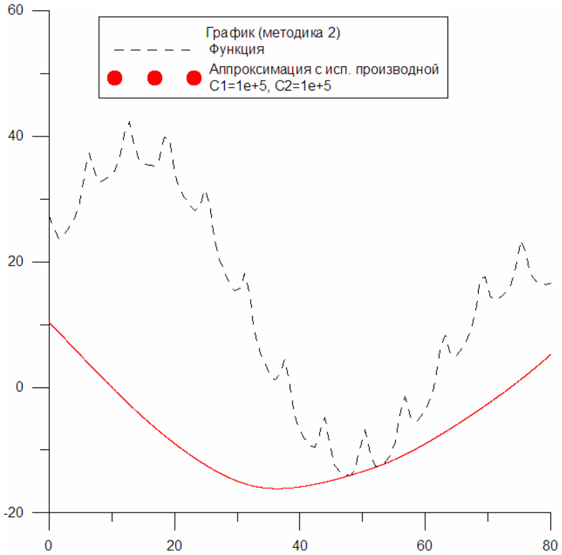
Методика овыпуклления



Методика овыпукления



Методика овыпукления



Методика овыпукления

- При применении методики овыпукления оператор Шепарада перестанет быть интерполянтом (становится аппроксимантом)

- При неограниченном увеличении C_1 и C_2, $F(x)$ стремится к стремится к прямой (одномерный случай) с углом наклона, соответствующем среднему значению производных в узлах

Методика восстановления производных

$$F(x) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x-x_i\|^4}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4}} \end{cases}$$

В узлах сетки градиенты аппроксимирующей функции - нулевые. Т.о. в каждом узле имеет место стационарная точка.

Методика восстановления производных

$$M_l(x) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i + a_i^l (x - x_i)}{\|x - x_i\|^4}}{\sum\limits_{i=1}^{k} \frac{1}{\|x - x_i\|^4}} \end{cases}$$

$$a_i^l = \frac{M_{l-1}(x + \Delta x) - M_{l-1}(x - \Delta x)}{2\Delta x}$$

$$a_i^0 = 0$$

Методика восстановления производных

1. Имеем набор $\{x_k, f_k\}$.
2. Полагаем $l = 0$. Строим M_0.
3. $l = l + 1$.
4. Вычисляем $a_k^l = \frac{M_{l-1}(x+\Delta x) - M_{l-1}(x-\Delta x)}{2\Delta x}$ из M_{l-1}.
5. Вычисляем M_l.
6. Вычисляем $J = \int (M_l(x) - M_{l-1}(x))^2 dx$.
7. Если $J < \varepsilon^*$, $M(x) = M_l$; конец процедуры.
8. Если $l > L^*$, $M(x) = M_l$; конец процедуры, иначе переход на пункт 2.

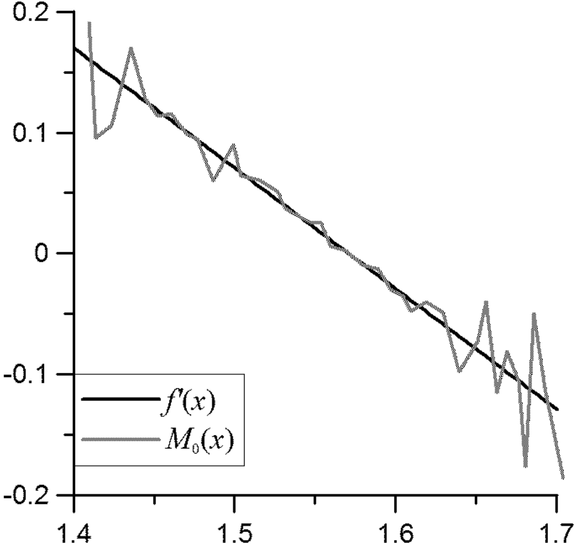
Методика восстановления производных

$$f(x) = sin(x)$$

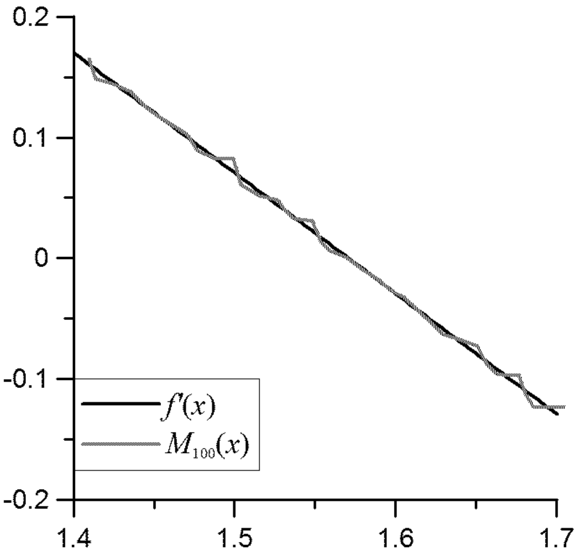
$$x \in [1.4, 1.8]$$

$$f'(x) = cos(x)$$

Методика восстановления производных

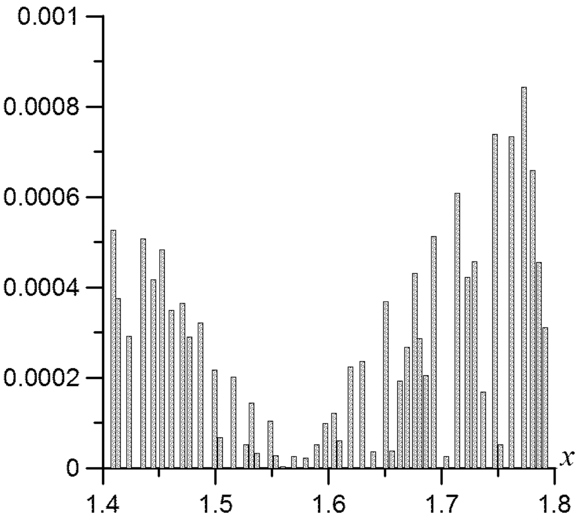


Методика восстановления производных



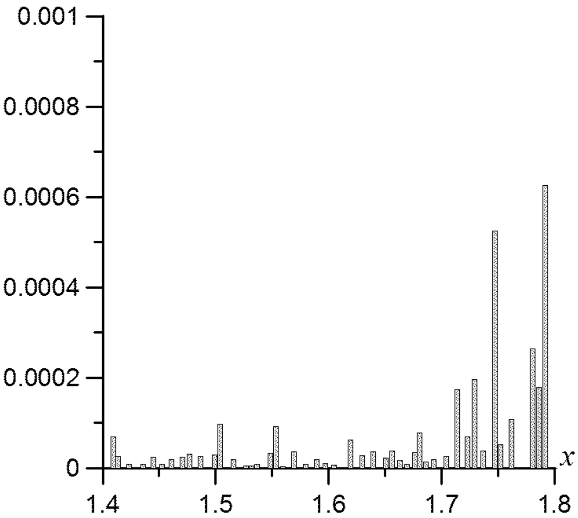
Методика восстановления производных

Гистограмма отклонения M_0 от $f(x)$



Методика восстановления производных

Гистограмма отклонения M_{100} от $f(x)$



Приложение 1
Arnold Neumaier

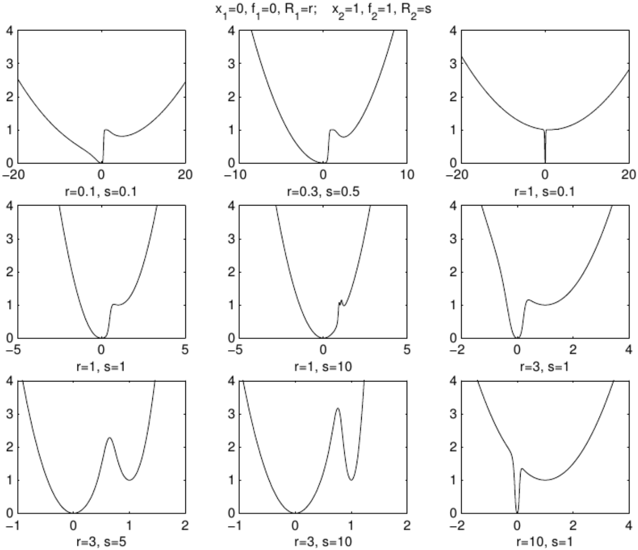
Arnold Neumaier. Rational Functions with Prescribed Global and Local Minimizers // Journal of Global Optimization. Volume 25 Issue 2, February 2003. Pages 175–181

Приложение 1. Arnold Neumaier

$$F(x) = \begin{cases} f_i & \text{если } x = x_i \\ \dfrac{\sum\limits_{i=1}^{k} (2f_i + r_i(x))/r_i(x)^2}{\sum\limits_{i=1}^{k} 2/r_i(x)^2} & \text{если } x \neq x_i \end{cases}$$

$$r_i(x) = \|R_i(x - x_i)\|^2$$

Приложение 1. Arnold Neumaier



Приложение 2
“Тяжелые” задачи

Приложение 2. Постановка задачи

$$f(x) \to min$$

$$\underline{x} \leq x \leq \overline{x}$$

- целевая функция невыпукла
- аналитические градиенты недоступны
- функция "вычисляется с шумами"
- время вычисления целевой функции "велико" (десятки минут, часы, ...)
- размерность "невелика" (2-5 переменных)
- требования к точности решения "небольшие"

Приложение 2. Источники “тяжелых” задач

- оптимизация “алгоритмически заданных функций”
- параметрическая идентификация динамических систем
- функция “вычисляется с шумами”
- оптимизация пучков траекторий
- задачи нормирования воздействий
- оптимизация стохастических систем
- анализ генных цепочек
- ...

Приложение 2. Требования к алгоритмам

- Нелокальная оптимизация
- Распараллеливание
- Устойчивость к “шумам”
- Высокая экономичность по числу вычислений функции

Приложение 2. Подходы

- База данных (все вычисленные значения)
- Локальные и “глобальные” алгоритмы
- Критерии остановки алгоритмов по числу проб
- “Когнитивная графика”
- Управляемый пакет запросов значения функции для параллельного устройства
- Model-Based Algorithm
- “Дневной” и “ночной” режимы
- Низкоуровневый диалоговый режим
- ...

Приложение 2. Аппроксимационые модели

- Модель Шепарда
- Радиально-базисные функции
- Сепарабельно-квадратичные
- Сплайн-функции
- ...

Приложение 2. Локальные автомат-алгоритмы

- Стохастический спуск
- Градиентный спуск
- Покоординатный спуск
- Сопряженные градиенты
- Метод Растригина
- Метод Пауэлла
- Локальный метод Шепарда
- Партан-методы
- ...

- 1D Шепард-графика
- 2D локальный спуск
- графический спуск по случайному направлению
- ...

Приложение 2. Глобальные автомат-алгоритмы

- Метод сеток
- Метод Лууса-Яаколы
- Метод "парабол"
- Сферический поиск
- Метод покрытий
- ...

Приложение 2. Глобальные диалог-алгоритмы

- 1D графика
- 2D графика
- Шепард-графика
- Метод “сгущений”
- ...

Приложение 2. Графическая подсистема

- 2D-проекции Шепард-аппроксимации
- 1D-проекции по случайному направлению
- “карта хребтов”
- “карта сгущений”
- “диаграммы базы проб” (1D, 2D)
- ...

Приложение 2. Аппроксимант Шепарда

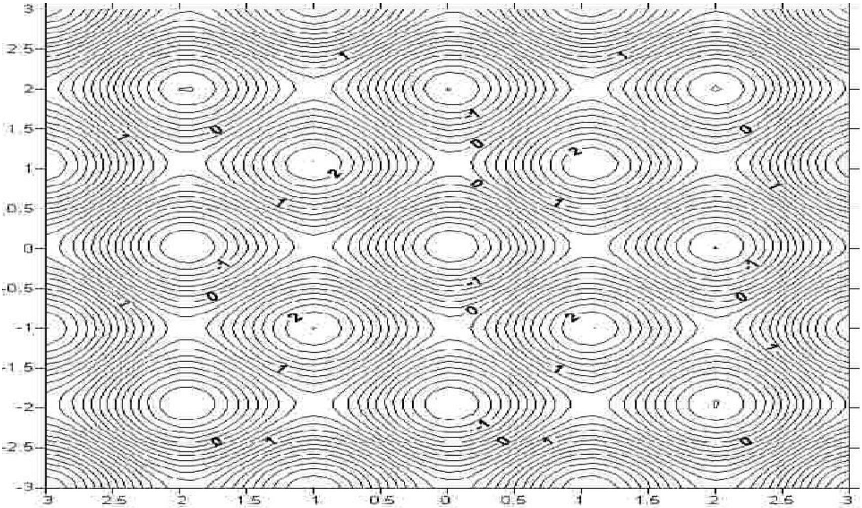
$$F(x) = \begin{cases} f_i \\ \\ \dfrac{\sum\limits_{i=1}^{k} \frac{f_i}{\|x-x_i\|^4}}{\sum\limits_{i=1}^{k} \frac{1}{\|x-x_i\|^4}} \end{cases}$$

Приложение 2. Тестовая задача

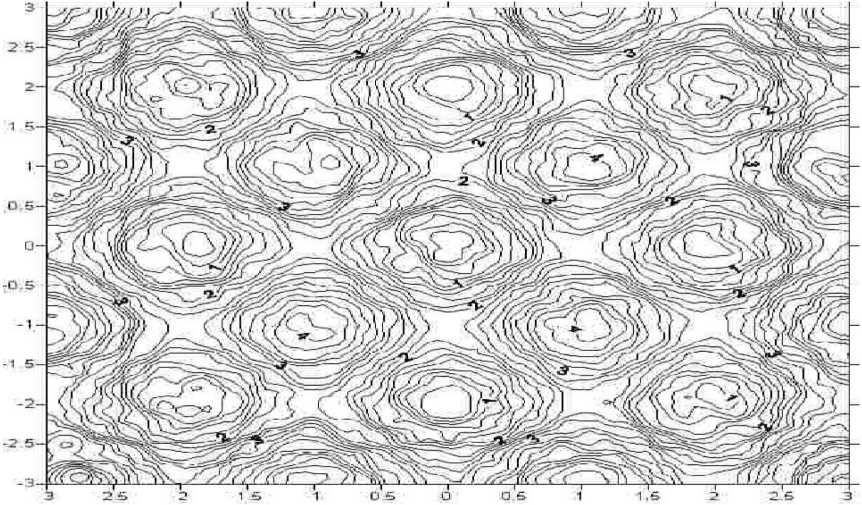
$$f(x) = \sum_i sin(\pi \cdot x_i - 0.5\pi) + 0.1 \sum_i x_i^2$$

$$-3 \leq x_i \leq 3$$
$$x^* = 0$$
$$f* = 0$$

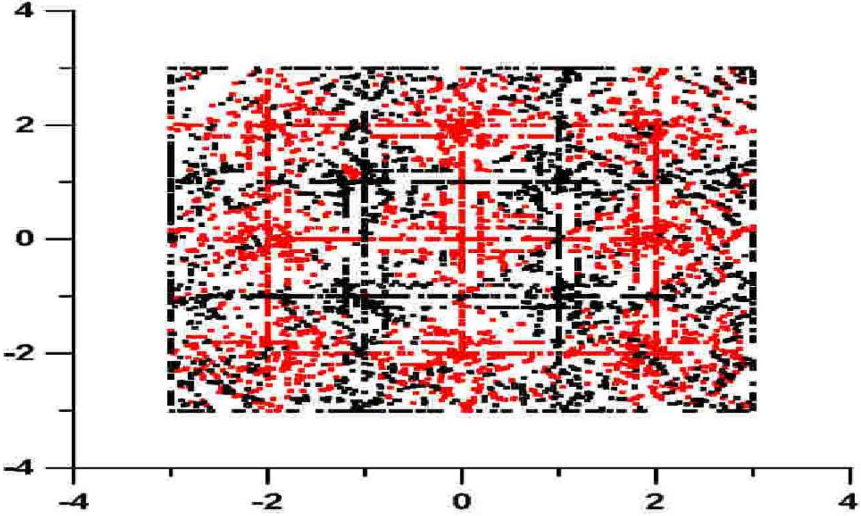
Приложение 2. Тест. Полная картина



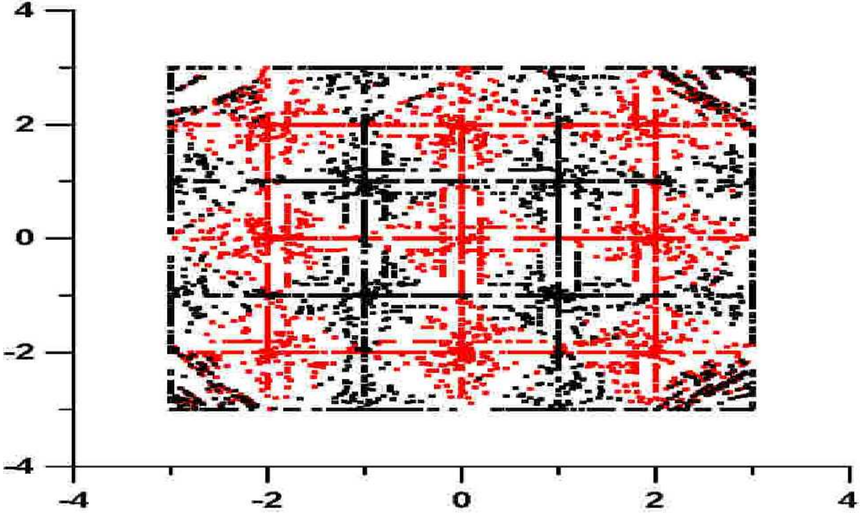
Приложение 2. Тест. 2D Шепард-проекция



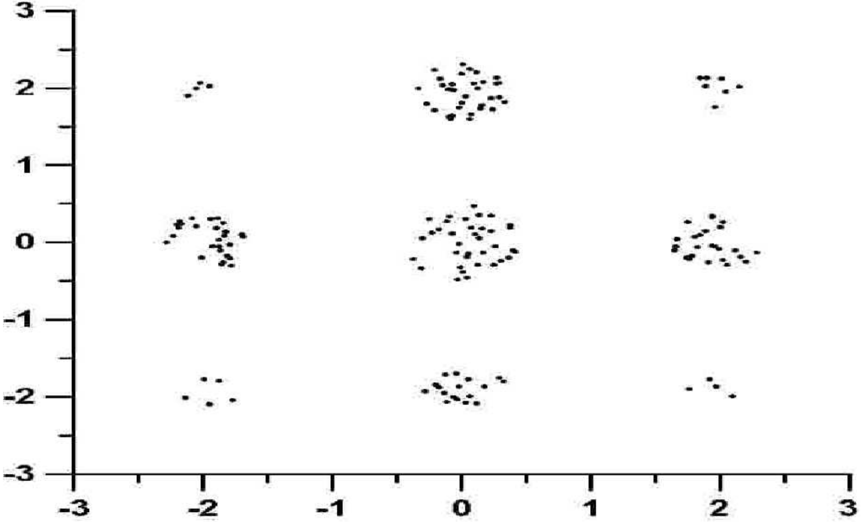
Приложение 2. Тест. Карта хребтов. 2000 проб



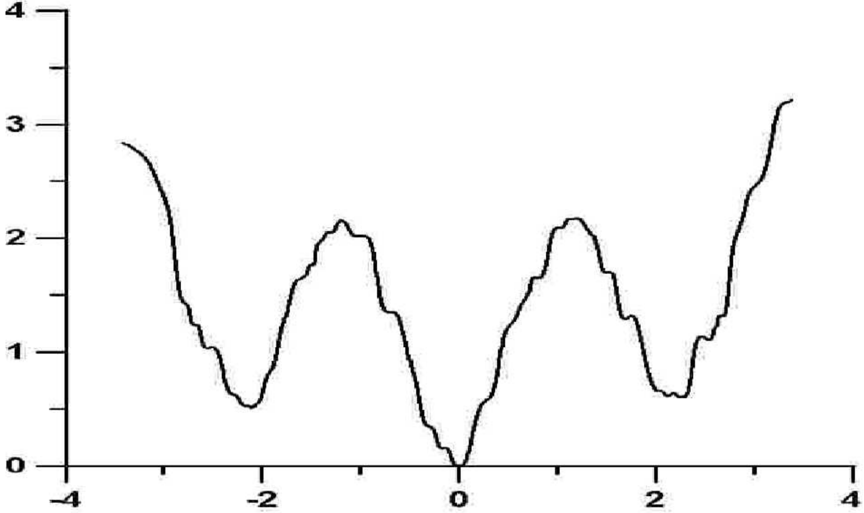
Приложение 2. Тест. Карта хребтов. 20000 проб



Приложение 2. Тест. Карта сгущений. 2000 проб



Приложение 2. Тест. Линейный поиск. 2000 проб



Приложение 2. “Широкий класс” задач

- Решение систем нелинейных уравнений
- Задачи математического программирования
- Релейные задачи оптимального управления
- Минимаксные задачи
- Двухуровневая оптимизация
- Оптимизация булевских функций
- ...

Приложение 3
Факторный анализ

факторный анализ – задача выявления существенных независимых переменных (показателей, факторов) неизвестной закономерности, представленной в виде измеренных экспериментальных данных

Задача прогнозирования женского бесплодия на основе данных анкетирования пациенток.

Постановщик – Кузьменко Е.Т. (Иркутский Диагностический Центр).

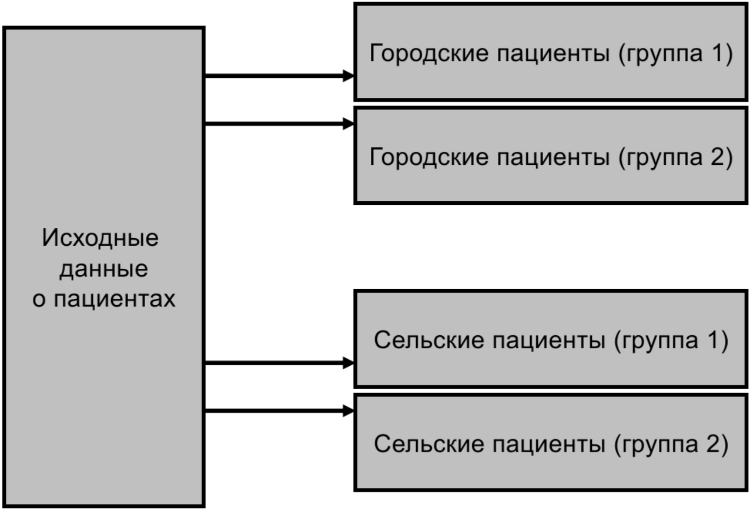
Исходные данные:

- число учитываемых факторов - 30
- данные опроса 320 пациенток

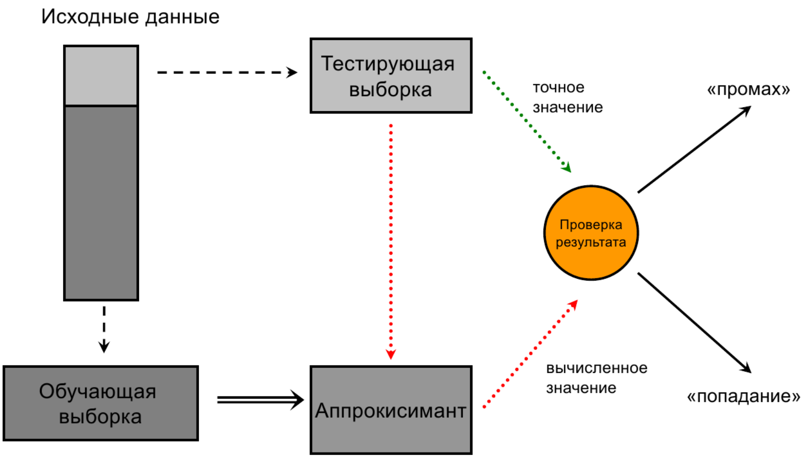
Желаемые результаты:

- создание модели для прогнозирования женского бесплодия;
- получение минимального набора наиболее значимых медико-экологических параметров (уменьшение размерности задачи).

Приложение 3. Разбиение исходных данных



Приложение 3. Метод комитетов



Погрешность полученной модели в процентах оценивается по формуле:

$$\Delta = \frac{C_E}{C_T} \cdot 100$$

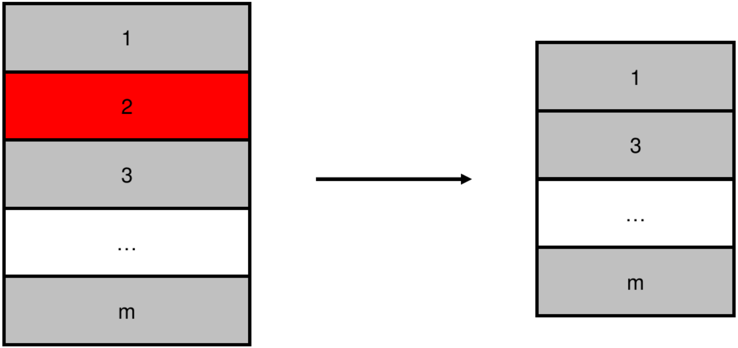
C_E – число "промахов" (неудачных точек аппроксимации)

C_T – число элементов обучающей выборки

Для уменьшения погрешности моделей предлагается следующая методика:

- “Горизонтальная чистка” - удаление точек входных данных, которые создают “шум”
- “Вертикальная чистка” - удаление факторов (переменных), которые создают “шум”

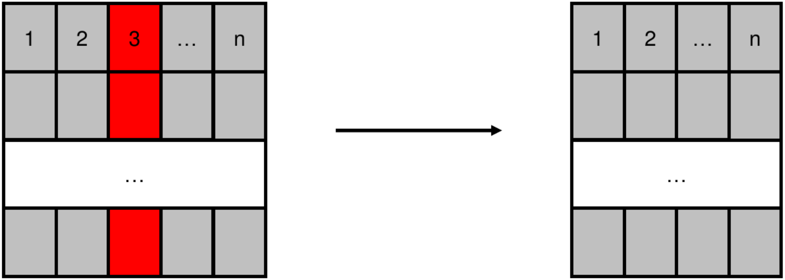
Приложение 3. “Горизонтальная чистка”



Приложение 3. “Горизонтальная чистка”

1. Разбиваем входные данные на обучающую (80%) и тестирующую (20%) выборки.
2. В цикле удаляем очередной элемент (точку) обучающей выборки, вычисляем разницу погрешности модели до и после удаления.
3. Повторяем пункты 1 и 2 достаточно большое число раз (1000), накапливая статистику.
4. Точку с наибольшей суммой погрешностей удаляем из входных данных.
5. Повторяем алгоритм нужное число раз.

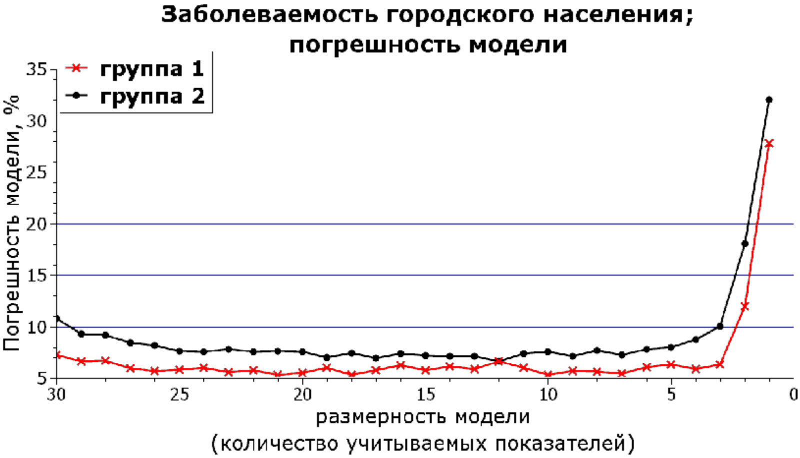
Приложение 3. “Вертикальная чистка”



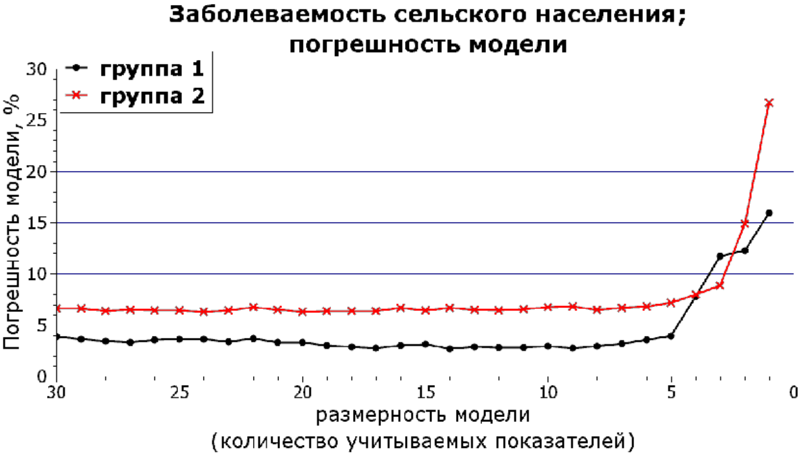
Приложение 3. “Вертикальная чистка”

1. Разбиваем входные данные на обучающую (80%) и тестирующую (20%) выборки.
2. В цикле удаляем очередной фактор (переменную), вычисляем разницу погрешности модели до и после удаления.
3. Повторяем пункты 1 и 2 достаточно большое число раз (1000), накапливая статистику.
4. Фактор (переменную) с наибольшей суммой погрешностей удаляем из входных данных.
5. Повторяем алгоритм нужное число раз.

Приложение 3. Погрешность полученной модели



Приложение 3. Погрешность полученной модели



Приложение 3. Выводы

1. Аппроксимант Шепарда может применяться при построении моделей на основе многомерных данных.
2. Предложенная методика "горизонтальной чистки" позволяет уменьшить погрешность модели за счет удаления данных, вносящих наибольший "шум".

Приложение 3. Выводы

3. Предложенная методика “вертикальной чистки” позволяет выявить наиболее значимые факторы и уменьшить размерность задачи.
4. Получено решение задачи факторного анализа медико-экологических данных. Найденное решение оценено экспертом как правдоподобное.

Публикации

- Shepard D. A two-dimensional interpolation function for irregularly-spaced data // Proc. of the 23 ACM National Conference, ACM Press, New York, 1968, pp. 517-524.
- Neumaier A. Rational functions with prescribed global and local minimizers // J. Global Optimization, 25 (2003), pp. 175-181.
- Caira R., Dell'Accio F. Shepard-Bernoulli operators // Mathematics of computation, 2007, V. 76, N. 257, pp. 299–321.

Публикации

- Горнов А.Ю., Кузьменко Е.Т., Аникин А.С., Зароднюк Т.С. Применение алгоритмов аппроксимации экспериментальных данных в задаче выявления значимых медико-социальных факторов // Современные технологии. Системный анализ. Моделирование. Иркутск: ИрГУПС, 2008. С. 92-96.
- Кузьменко В.В., Аникин А.С. Методика статического моделирования по Шепарду для оценки концентрации фракций холестерина в плазме // Ученые заметки ТОГУ: Электрон. науч. изд. 2010. Т. 1, N 1. С. 59-61.
- Кузьменко В.В., Аникин А.С., Горнов А.Ю., Мирошниченко И.А. Алгоритм оценки концентрации холестерина липопротеинов с использованием методики статического моделирования по Шепарду // Труды XV Байкальской Всероссийской конференции “Информационные и математические технологии в науке и управлении”. Часть I. Иркутск: ИСЭМ СО РАН, 2010. С. 145-151.

Публикации

- A.Yu. Gornov, I.A. Veyalko. The global extremum searching algorithm for bang-bang optimal control problem based on the Shepard operator // J. Studia Informatica Universalis. 2011. Vol. 9. Num. 3 pp. 91-104.
- Горнов А.Ю., Финкельштейн Е.А. Метод Шепарда аппроксимации таблично заданных функций и его обобщения // Доклады 9-ой международной конференции "Интеллектуализация обработки информации". М.: Торус Пресс, 2012. С. 98-101.
- Зароднюк Т.С. Алгоритм поиска глобального экстремума с использованием семейства функций Шепарда // Сборник научных трудов Всероссийской конференции с международным участием "Информационные и математические технологии в науке, технике, медицине", Томск, 5-8 ноября 2012 г. Томск: Издательство ТПУ, 2012. Т. 1. С. 49-51.

Публикации

- Горнов А.Ю., Веялко И.А. Метод Шепарда для решения релейных задач оптимального управления // Труды XVII Байкальской всероссийской конференции “Информационные и математические технологии в науке и управлении”. Иркутск-Байкал, 30 июня-9 июля 2012. Иркутск: ИСЭМ СО РАН. Т. 3. С. 218-224.
- Доржиева А.Б. Алгоритм построения аппроксимации многоэкстремальной функции на основе метода Шепарда // Труды XVIII Байкальской всероссийской конференции “Информационные и математические технологии в науке и управлении”. Часть III. Иркутск: ИСЭМ СО РАН, 2013. С. 238-241.

МЕТОД АППРОКСИМАЦИИ ШЕПАРДА И ЕГО ПРИЛОЖЕНИЯ

. . .

SHEPARD APPROXIMATION METHOD AND IT'S APPLICATIONS

Горнов А.Ю. gornov@icc.ru

ИДСТУ СО РАН, Иркутск