

Суррогатное моделирование и оптимизация профиля крыла самолета на основе гауссовских процессов

Евгений Бурнаев, Павел Ерофеев, Алексей Зайцев, Даниил Кононенко, Ермак Капушев

1. Институт Проблем Передачи Информации,
127994, г.Москва, ГСП-4, Большой Каретный переулок, 19, стр.1

2. DATADVANCE,
105064, г.Москва, Садово-Черногрязская улица, 13/3

3. МФТИ, PreMoLab
141700, г.Долгопрудный, Институтский переулок, 9

{evgeny.burnaev,pavel.erofeev,alexey.zaytsev,daniil.kononenko,ermek.kapushv}@datadvance.net

Аннотация

В работе рассматривается задача оптимизации ресурсоемкой функции. Кроме самой целевой (точной) функции задано также некоторое ее приближение — грубая функция. Для решения оптимизационной задачи предлагается использовать суррогатное моделирование. В данной работе суррогатная модель строится на основе гауссовских процессов, при этом используются значения как точной, так и грубой функции. Результаты оптимизации на основе такой модели сравниваются с результатами прямой оптимизации и с оптимизацией на основе суррогатных моделей, использующих только точную функцию. Работа алгоритмов продемонстрирована на задаче оптимизации профиля крыла самолета.

1. Введение

Одной из задач, возникающих при проектировании самолета является оптимизация профиля крыла с целью увеличения аэродинамического качества — отношения коэффициента подъемной силы к коэффициенту лобового сопротивления [1]. Для вычисления аэродинамических коэффициентов используются CFD-коды [1, 2]. CFD-коды, предсказывающие значения аэродинамических коэффициентов с большой точностью, требуют значительных вычислительных ресурсов [3]. Поэтому при решении задачи оптимизации требуется минимизировать количество вызовов CFD-кода. Для этого предлагается использовать суррогатное моделирование. Кроме того, для сокращения времени решения оптимизационной задачи возможно использование более грубой, но быстрой модели для вычисления аэродинамических

коэффициентов, поскольку это позволит строить более точные суррогатные модели, используя меньше вызовов точной функции.

В данной работе рассматриваются методы построения суррогатных моделей на основе гауссовских процессов и их применение в задаче оптимизации. Рассмотренные подходы демонстрируются на реальной задаче оптимизации профиля крыла самолета.

Статья устроена следующим образом:

- в разделе 2 описывается постановка задачи оптимизации;
- в разделе 3 описываются методы построения суррогатных моделей на основе гауссовских процессов;
- в разделе 4 рассказывается о методах оптимизации на основе суррогатных моделей;
- в разделе 5 описывается использованная в работе параметризация профиля крыла самолета;
- в разделах 6 и 7 приводятся результаты работы рассмотренных подходов на реальных данных, проводится их сравнение с другими алгоритмами оптимизации.

2. Постановка задачи

Сформулируем задачу оптимизации профиля крыла самолета. Пусть $C_l(\mathbf{x}, \alpha, M)$ и $C_d(\mathbf{x}, \alpha, M)$ аэродинамические коэффициенты подъемной силы и сопротивления соответственно, $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^{59}$ — вектор 59-мерного пространства, описывающий профиль крыла самолета, α — угол атаки, M — число Маха. Отношение $\frac{C_l(\mathbf{x}, \alpha, M)}{C_d(\mathbf{x}, \alpha, M)}$ называется аэроди-

намическим качеством. Необходимо решить следующую задачу:

$$\max_{\mathbf{x}, \alpha} \frac{C_l(\mathbf{x}, \alpha, M)}{C_d(\mathbf{x}, \alpha, M)} \text{ s.t. } \begin{cases} \alpha^{\min} \leq \alpha \leq \alpha^{\max}, \\ t_{fs}(\mathbf{x}) \geq t_{fs}^0, \\ t_{rs}(\mathbf{x}) \geq t_{rs}^0, \end{cases} \quad (1)$$

где ε , α^{\min} , α^{\max} , t_{rs}^0 , t_{fs}^0 — заданные константы, число Маха M — фиксировано. В (1) последние два неравенства выражают ограничение на толщину крыла (переднего и заднего лонжерона соответственно). $C_l(\mathbf{x}, \alpha, M)$ и $C_d(\mathbf{x}, \alpha, M)$ — заданы в виде “черного ящика” (CFD-кода). Такой CFD-код будем называть точной моделью или точной функцией. Задано также грубое приближение точной модели (например, тот же самый CFD-код с более грубой сеткой). Вычисление такого приближения требует меньше вычислительных ресурсов.

3. Суррогатное моделирование

Пусть задана обучающая выборка $D = (X, Y) = \{(\mathbf{x}_i, y_i = f(\mathbf{x}_i)), i = 1, \dots, N\}$, где $y = f(\mathbf{x})$ некоторая неизвестная функция, $\mathbf{x} \in X \subset \mathbb{R}^n$, $y \in \mathbb{R}$. Необходимо по заданным значениям восстановить неизвестную зависимость $y = f(\mathbf{x})$.

3.1. Гауссовские процессы

Одним из способов построения суррогатных моделей является регрессия на основе гауссовских процессов. Будем считать, что $f(\mathbf{x})$ является реализацией гауссовского процесса. Такие процессы полностью определяются своей функцией среднего $\mu(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}))$ и ковариационной функцией $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$. О виде ковариационной функции делается априорное предположение. Обычно предполагают, что она принадлежит экспоненциальному семейству:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\sum_{i=1}^n \theta_i |x_i - x'_i|^2\right) + \sigma_n^2 \delta(\mathbf{x}, \mathbf{x}'), \quad (2)$$

где σ_n^2 — регуляризирующий параметр шума. Положив математическое ожидание $\mu(\mathbf{x})$ равным нулю, запишем логарифм правдоподобия [8]:

$$-\frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \ln |\mathbf{K}| - \frac{Y^T \mathbf{K}^{-1} Y}{2\sigma^2} \quad (3)$$

где $\mathbf{K} = K(X, X) = \|k(\mathbf{x}_i, \mathbf{x}_j)\|_{i,j=1}^N$, $|\mathbf{K}|$ — детерминант матрицы \mathbf{K} . Оценки параметров θ и σ^2 получаются максимизацией (3). Используя найденные оценки $\hat{\theta}$ и $\hat{\sigma}^2$, апостериорная оценка математического ожидания гауссовского процесса запишется в следующем виде [7]:

$$\hat{f}(\mathbf{x}) = \mathbf{k}^T \mathbf{K}^{-1} Y, \quad (4)$$

где $\mathbf{k} = [k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_N)]^T$. При этом апостериорная ковариационная функция имеет вид:

$$s^2(\mathbf{x}) = \mathbb{V}[\hat{f}(\mathbf{x})] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \quad (5)$$

Дисперсия гауссовского процесса может быть использована в качестве оценки ошибки аппроксимации в заданной точке.

3.2. Разноточные гауссовские процессы

Пусть кроме самой функции $f(\mathbf{x})$, которую теперь будем обозначать $f_h(\mathbf{x})$ и называть точной функцией, задано некоторое ее приближение $f_l(\mathbf{x})$ — грубая функция. Заданы две выборки: $D_h = (X_h, Y_h) = \{(\mathbf{x}_i^h, y_i^h), y_i^h = f_h(\mathbf{x}_i^h), i = 1, \dots, N_h\}$, $D_l = (X_l, Y_l) = \{(\mathbf{x}_i^l, y_i^l), y_i^l = f_l(\mathbf{x}_i^l), i = 1, \dots, N_l\}$. По заданным значениям необходимо построить аппроксимацию точной функции $f_h(\mathbf{x})$. Введем следующие обозначения:

$$X = \begin{pmatrix} X_h \\ X_l \end{pmatrix}, \quad Y = \begin{pmatrix} Y_h \\ Y_l \end{pmatrix}.$$

Также, как и в предыдущем разделе, значение в точке \mathbf{x} рассматривается как реализация гауссовского процесса. Будем использовать модель, введенную в статье [5], которая предполагает, что

$$\text{cov}[f_h(\mathbf{x}_i), f_l(\mathbf{x}) | f_l(\mathbf{x}_i)] = 0, \quad \forall \mathbf{x} \neq \mathbf{x}_i. \quad (6)$$

Это означает, что если известно $f_l(\mathbf{x}_i)$, то из значения грубой функции в любой другой точке \mathbf{x} нельзя извлечь больше информации о значении точной функции в точке \mathbf{x}_i . Точную модель будем аппроксимировать суммой грубой модели $f_l(\mathbf{x})$, умноженной на масштабирующий коэффициент ρ , и некоторого независимого от $f_l(\mathbf{x})$ гауссовского процесса $f_d(\mathbf{x})$:

$$f_h(\mathbf{x}) = \rho f_l(\mathbf{x}) + f_d(\mathbf{x}), \quad (7)$$

Для такой модели выполняется (6). При этом получаем следующую ковариационную матрицу [6]:

$$\mathbf{K} = \begin{pmatrix} K_l(X_l, X_l) & \rho K_l(X_l, X_h) \\ \rho K_l(X_h, X_l) & \rho^2 K_l(X_h, X_h) + K_d(X_h, X_h) \end{pmatrix} \quad (8)$$

Здесь $K_l(X_l, X_l) = \|k_l(\mathbf{x}_i^l, \mathbf{x}_j^l)\|_{i,j=1}^{N_l}$, $K_l(X_l, X_h) = \|k_l(\mathbf{x}_i^l, \mathbf{x}_j^h)\|_{i=1, j=1}^{N_l, N_h}$, $K_d(X_h, X_h) = \|k_d(\mathbf{x}_i^h, \mathbf{x}_j^h)\|_{i,j=1}^{N_h}$, где $k_l(\mathbf{x}, \mathbf{x}')$ и $k_d(\mathbf{x}, \mathbf{x}')$ — ковариационные функции гауссовских процессов f_l и f_d соответственно. Параметры $f_l(\mathbf{x})$ настраиваются по выборке D_l , используя процедуру, описанную в предыдущем разделе. Обозначим через Y_d значения f_d в точках X_h :

$$Y_d = Y_h - \rho f_l(X_h).$$

Если значения грубой функции в точках X_h не заданы, то используется оценка $\hat{f}_l(X_h)$. Оценки параметров θ_d , σ_d^2 , процесса f_d и параметра ρ находятся

максимизацией логарифма правдоподобия:

$$-\frac{N_h}{2} \ln(\sigma_d^2) - \frac{1}{2} \ln |K_d(X_h, X_h)| - \frac{Y_d^T K_d(X_h, X_h)^{-1} Y_d}{2\sigma_d^2}. \quad (9)$$

Оценка значения точной функции в точке \mathbf{x} [6] имеет вид:

$$\hat{f}_h(\mathbf{x}) = \mathbf{k}^T \mathbf{K}^{-1} Y, \quad (10)$$

где

$$\mathbf{k} = \begin{pmatrix} \rho k_l(\mathbf{x}, X_l) \\ \rho^2 k_l(\mathbf{x}, X_h) + k_d(\mathbf{x}, X_h) \end{pmatrix}$$

Оценка дисперсии в точке \mathbf{x} [6]:

$$s^2(\mathbf{x}) = \mathbb{V}[\hat{f}_h(\mathbf{x})] = \hat{\rho}^2 k_l(\mathbf{x}, \mathbf{x}) + k_d(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \quad (11)$$

Отметим, что при малых значениях регуляризующего параметра (см. (2)) полученная оценка является интерполирующей, то есть, если \mathbf{x} совпадает с точкой обучающей выборки точной функции $\mathbf{x} = \mathbf{x}_i^h$, то: $\mathbb{V}[\hat{f}_h(\mathbf{x}_i^h)] = 0$, $\hat{f}_h(\mathbf{x}_i^h) = f_h(\mathbf{x}_i^h)$.

4. Оптимизация на основе суррогатных моделей

Один из возможных способов оптимизации на основе суррогатных моделей — это поиск минимума аппроксиматора, построенного по какой-то начальной обучающей выборке. Однако такой алгоритм может привести к локальному минимуму [9]. При построении суррогатной модели может оказаться так, что в обучающей выборке нет точек, лежащих в области глобального минимума, однако есть точки, находящиеся возле локального минимума. При этом значение, предсказываемое суррогатной моделью в точке глобального минимума, может получиться больше предсказания в локальном минимуме. В этом случае алгоритм находит локальный минимум. Это объясняется тем, что при таком подходе не учитывается ошибка (дисперсия) предсказываемых значений, то есть в области с большой дисперсией значение функции может оказаться меньше значения, предсказываемого аппроксиматором. В работе используется алгоритм, учитывающий это, и позволяющий найти глобальный минимум.

4.1. Ожидаемое улучшение

Пусть $y_{min} = \min(y_1, \dots, y_n)$ — текущее наилучшее значение целевой функции. Точно значение функции в произвольной точке \mathbf{x} неизвестно до тех пор, пока оно не будет вычислено. Такую ситуацию можно смоделировать представив значение функции в точке \mathbf{x} случайной нормально распределенной величиной $Y(\mathbf{x})$ с математическим ожиданием и дисперсией, равными математическому ожиданию и дисперсии соответствующей суррогатной модели.

Так как $Y(\mathbf{x})$ случайная величина, то существует вероятность, что ее значение будет меньше, чем y_{min} . Улучшение значения в точке \mathbf{x} есть $I(\mathbf{x}) = \max(y_{min} - Y(\mathbf{x}), 0)$. Величина

$$\mathbb{E}[I(\mathbf{x})] = \mathbb{E}[\max(y_{min} - Y(\mathbf{x}), 0)]$$

называется ожидаемым улучшением. Она выражает наше ожидание относительно того, насколько значение в точке \mathbf{x} окажется меньше y_{min} . Поскольку $Y(\mathbf{x})$ нормально распределенная случайная величина, то $\mathbb{E}[I(\mathbf{x})]$ имеет вид:

$$\mathbb{E}[I(\mathbf{x})] = \begin{cases} (y_{min} - \hat{f}(\mathbf{x}))\Phi\left(\frac{y_{min} - \hat{f}(\mathbf{x})}{s(\mathbf{x})}\right) + s(\mathbf{x})\phi\left(\frac{y_{min} - \hat{f}(\mathbf{x})}{s(\mathbf{x})}\right), & \text{если } s(\mathbf{x}) > 0 \\ 0, & \text{если } s(\mathbf{x}) = 0. \end{cases} \quad (12)$$

В приведенной выше формуле $\phi(\cdot)$ и $\Phi(\cdot)$ — плотность вероятности и функция стандартного нормального распределения.

Для поиска минимума применяется алгоритм Эффективной Глобальной Оптимизации (EGO) [9]:

Шаг 1. Сгенерировать начальную обучающую выборку.

Шаг 2. По обучающей выборке построить суррогатную модель.

Шаг 3. Найти точку, доставляющую максимум функции $\mathbb{E}[I(\mathbf{x})]$:

$$\mathbf{x}^* = \operatorname{argmax} \mathbb{E}[I(\mathbf{x})]$$

Шаг 4. Вычислить $y^* = f(\mathbf{x}^*)$.

Шаг 5. Если не выполнен критерий останова, то добавить (\mathbf{x}^*, y^*) к обучающей выборке, выполнить шаги 2-5.

Заметим, что $\mathbb{E}[I(\mathbf{x})] = 0$ при $s(\mathbf{x}) = 0$, поэтому в случае интерполирующего аппроксиматора новая точка не будет совпадать ни с одной из точек выборки. Данный факт гарантирует сходимость метода, то есть нахождение глобального минимума с ростом размера выборки [10].

5. Параметризация профиля крыла самолета

Рассматривается двумерный профиль крыла самолета. Он достаточно полно описывается вектором 59-мерного пространства $\mathbf{x} \in \mathbb{R}^{59}$. Профиль крыла самолета представляет собой замкнутую кривую. Для ее параметризации используют набор из 59-и точек (x, y) . Ось x направлена вдоль хорды профиля крыла, на ней выбрано 30 точек: $x_1 = 1$, $x_{30} = 0$, остальные точки лежат между 0 и 1 в порядке убывания.

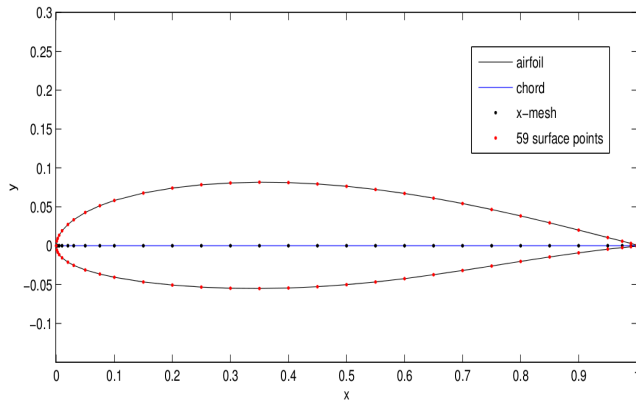


Рис. 1. Параметризация профиля крыла

В каждой точке x_i задается y_i и y_{30+i} , $i = 1, \dots, 29$, $y_{30} = 0$, описывающие верхнюю и нижнюю части профиля крыла соответственно. Точки x фиксированы, и профили крыла самолета отличаются только y компонентами, поэтому каждый профиль описывается 59-мерным вектором $\{y_k\}_{k=1}^{59}$. Такая параметризация имеет 2 существенных недостатка. Во-первых, она имеет очень большую размерность. Во-вторых, в исходном 59-мерном пространстве сложно описать допустимое множество точек, соответствующих физическим профилям крыла самолета.

Чтобы преодолеть эти трудности, в данной работе применяется снижение размерности методом главных компонент [11]. Для нахождения главных компонент использовалась база из 206 различных корректных профилей крыла самолета. Опыты показывают, что для описания профилей с хорошей точностью достаточно использовать первые девять компонент. Такая параметризация охватывает достаточно широкий класс различных профилей и при этом использует небольшое число параметров.

6. Экспериментальные результаты

В ходе эксперимента сравнивалась работа трех алгоритмов глобальной оптимизации, а именно:

- алгоритм глобальной оптимизации DIRECT, описанный в статье [13];
- алгоритм EGO с использованием суррогатной модели на основе гауссовских процессов;
- алгоритм EGO с использованием суррогатной модели на основе разноточных гауссовских процессов.

В качестве точной и грубой функции использовались 2 различных CFD-кода. Время вычисления значений C_l и C_d в одной точке составляло 12-15 ми-

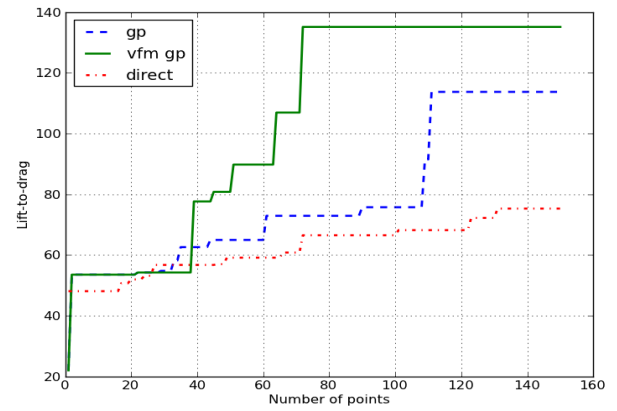


Рис. 2. Зависимость найденного оптимального значения от количества вызовов точной функции.

нут для точной функции и ~ 2 секунды для грубой функции. Для построения суррогатных моделей была сгенерирована выборка точных данных размера $N_h = 25$ и выборка грубых данных размера $N_l = 170$. В данном эксперименте сравнивались максимальные значения, найденные сравниваемыми алгоритмами при одинаковом количестве вызовов точной функции. Результаты численного эксперимента приведены на рисунке 2. По значениям C_l и C_d было построено множество Парето (рисунок 3). На графике видно, что множество Парето полученное при помощи алгоритмов на основе суррогатных моделей лежит ниже множества, построенного алгоритмом прямой оптимизации. Заметим, что алгоритмы оптимизации на основе суррогатных моделей значительно превосходят алгоритм прямой оптимизации, а использование данных разного качества позволяет получить более точное решение используя такое же количество вызовов точной функции.

7. Заключение

В работе была продемонстрирована эффективность применения суррогатных моделей на основе разноточных гауссовских процессов в задаче глобальной оптимизации. Использование грубой функции позволяет строить более точные суррогатные модели при меньшем количестве вызовов ресурсоемкой точной функции. Таким образом, применение таких моделей снижает время решения оптимизационной задачи. Данная работа является подготовительным этапом к решению реальной задачи оптимизации профиля крыла самолета. При проектировании самолета оптимизацию профиля крыла проводят при дополнительных нелинейных ограничениях на аэродинамические коэффициенты, вычисление которых требует значительных вычислитель-

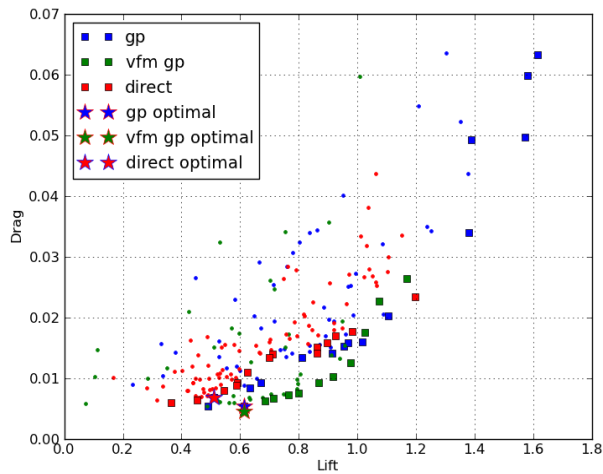


Рис. 3. Множество оптимальных по Парето решений.

ных ресурсов. Дальнейшие исследования предполагается сосредоточить на решении этой задачи. Работа выполнена при поддержке Лаборатории структурных методов анализа данных в предсказательном моделировании, МФТИ, грант правительства РФ дог. 11.G34.31.0073

Список литературы

- [1] Besnard E., Schmitz A., Boscher E., Garcia N., Cebeci T. Two-dimensional Aircraft High Lift Ssystem Design and Optimization. // AIAA paper 98-0123, 1998.
- [2] Blessing B.H., Pham J., Marshall D.D. Using CFD as a Desing Tool on New Innovative Airliner Configurations. // 47th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, January 5, 2009, pp. 1-24.
- [3] Jameson A., Martinelli L., Vassberg J. C. Using Computational Fluid Dynamics for Aerodynamics — A Critical Assessment // 23rd International Congress of Aeronautical Sciences, September 8-13, 2002
- [4] Forrester A., Sobester A., Keane A. Engineering Design via Surrogate Modelling. A Practical Guide. – Wiley, 2008. – 238 p.
- [5] Kennedy M.C., O’Hagan A. Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. // Biometrika, Vol. 87, No. 1. (Mar., 2000), pp. 1-13
- [6] Forrester A.I.J., Sobester A., Keane A.J. Multi-fidelity optimization via surrogate modelling. // Computational Engineering and Design Group, School of Engineering Sciences, University of Southampton, Southampton SO17 1BJ, UK
- [7] Rasmussen C.E., Williams C.K.I. Gaussian Processes for Machine Learning. – the MIT Press, 2006.
- [8] Jones D.R. A taxonomy of global optimization methods based on response surfaces. // Journal of Global

Optimization, 21:345–383, 2001.

- [9] Jones D.R., Schonlau M., Welch W.J. Efficient Global Optimization of Expensive Black-Box Functions. // Journal of Global Optimization 13: 455–492, 1998
- [10] Locatelli M. Bayesian algorithms for one-dimensional global optimization. // Journal of Global Optimization, 10(1):57–76, January 1997.
- [11] Jolliffe I.T. Principal Component Analysis. // Springer, second edition, 2002.
- [12] Schonlau M. Computer Experiments and Global Optimization // Ph.D. Thesis, University of Waterloo, 1997.
- [13] Jones D.R. DIRECT Global Optimization Algorithm // Encyclopedia of Optimization, Kluwer Academic Publishers, 2001