КАК РЕШАТЬ ЗАДАЧИ, КОТОРЫЕ НЕ РЕШАЮТСЯ НА СУПЕРКОМПЬЮТЕРАХ?

Евгений Евгеньевич Тыртышников

Институт вычислительной математики Российской академии наук Московский государственный университет им. М.В.Ломоносова Московский физико-технический институт eugene.tyrtyshnikov@gmail.com

Москва, 22 ноября 2016



РАЗМЕРНОСТЬ – ЭТО ПРОКЛЯТИЕ ИЛИ БЛАГО?

Как задать элементы d-мерной матрицы (тензора) $A = [a(i_1, \ldots, i_d)]$ размера $n \times \ldots \times n$?

▶ Если d=300 и n=2, то число этих элементов $2^{300}\gg 10^{83}$ больше числа атомов во вселенной!



СПЕЦИАЛЬНЫЕ ПРЕДСТАВЛЕНИЯ ТЕНЗОРОВ

Чистый тензор (тензор ранга 1, скелетон):

$$a(i,j,k) = u(i)v(j)w(k).$$

Разделение переменных.

Каноническое разложение – это сумма чистых тензоров:

$$a(i,j,k) = \sum_{\alpha=1}^{r} u(i,\alpha)v(j,\alpha)w(k,\alpha)$$

Определяется тремя матрицами: U = [u(:,1),...,u(:,r)],

$$V = [v(:,1),...,v(:,r)], W = [w(:,1),...,w(:,r)].$$



КАНОНИЧЕСКОЕ РАЗЛОЖЕНИЕ

- ▶ Применяется как модель для данных. Флюоресцентная спектрометрия: n_1 смесей, найти число веществ и их концентрации. Данные измерений = массив размеров $n_1 \times n_2 \times n_3$, n_2 и n_3 для частот излучателей и приемников.
 - Каноническое разложение дает число веществ и концентрации.
- ▶ Применяется в теории сложности основное средство при вычислении билинейных форм (Strassen, Pan, Bini и др.).
- Много трудных проблем теоретических и вычислительных!

тензоры ранга < r \rightarrow тензор ранга r



УМНОЖЕНИЕ 2×2 -МАТРИЦ

Правило "строка на столбец" дает 8 умножений:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

АЛГОРИТМ ШТРАССЕНА ДЛЯ БЫСТРОГО УМНОЖЕНИЯ МАТРИЦ

В 1969 Штрассен (Strassen) открыл алгоритм, в котором всего 7 умножений. При умножении блочных 2×2 -матриц алгоритм Штрассена содержит 7 умножений блоков:

$$\alpha_{1} = (a_{11} + a_{22})(b_{11} + b_{22})
\alpha_{2} = (a_{21} + a_{22})b_{11}
\alpha_{3} = a_{11}(b_{12} - b_{22})
\alpha_{4} = a_{22}(b_{21} - b_{11})
\alpha_{5} = (a_{11} + a_{12})b_{22}
\alpha_{6} = (a_{21} - a_{11})(b_{11} + b_{12})
\alpha_{7} = (a_{12} - a_{22})(b_{21} + b_{22})$$

$$c_{11} = \alpha_{1} + \alpha_{4} - \alpha_{5} + \alpha_{7}
c_{12} = \alpha_{3} + \alpha_{5}
c_{21} = \alpha_{2} + \alpha_{4}
c_{22} = \alpha_{1} + \alpha_{3} - \alpha_{2} + \alpha_{6}$$

КАНОНИЧЕСКОЕ ТЕНЗОРНОЕ РАЗЛОЖЕНИЕ ДЛЯ ПОИСКА БЫСТРОГО АЛГОРИТМА УМНОЖЕНИЯ $n \times n$ МАТРИЦ

$$\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} \quad c_k = \sum_{i=1}^{n^2} \sum_{j=1}^{n^2} h_{ijk} \ a_i \ b_j$$

$$\mathbf{h}_{ijk} = \sum_{\alpha=1}^{R} \mathbf{u}_{i\alpha} \ \mathbf{v}_{j\alpha} \ \mathbf{w}_{k\alpha}$$

$$\Rightarrow c_k = \sum_{\alpha=1}^{R} \mathbf{w}_{k\alpha} \left(\sum_{i=1}^{n^2} u_{i\alpha} a_i \right) \left(\sum_{j=1}^{n^2} v_{j\alpha} b_j \right)$$

Теперь лишь *R* умножений блоков!

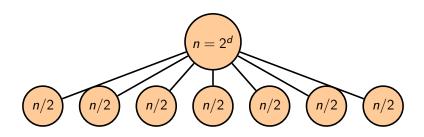
Если n = 2, то R = 7 (Strassen, 1969).

Рекурсия \Rightarrow $O(n^{\log_2 7})$ умножений для произвольного n.

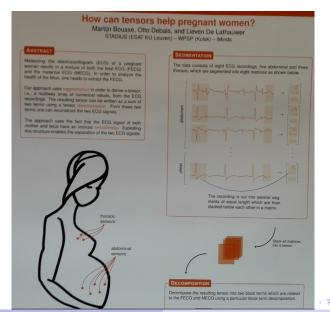


РЕКУРСИЯ ДЛЯ БЫСТРОГО УМНОЖЕНИЯ МАТРИЦ

Используя рекурсию, можно перемножить две матрицы порядка $n=2^d$ с затратой не более $7^d=n^{\log_27}$ и $7n^{\log_27}$ сложений-вычитаний.



ТЕНЗОРЫ ПОВСЮДУ



УВИДЕТЬ В ТЕНЗОРАХ МАТРИЦЫ

$$a(i_1; i_2, i_3) = \sum g_1(i_1, \alpha_1) a_1(\alpha_1; i_2, i_3)$$

$$a_1(\alpha_1, i_2; i_3) = \sum g_2(\alpha_1, i_2; \alpha_2)g_3(\alpha_2; i_3)$$

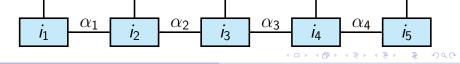
$$a(i_1, i_2, i_3) = \sum g_1(i_1, \alpha_1)g_2(\alpha_1, i_2, \alpha_2)g_3(\alpha_2, i_3)$$

TEH3OPHЫЙ ПОЕЗД = MATRIX PRODUCT STATE= ЛИНЕЙНАЯ ТЕНЗОРНАЯ СЕТЬ

$$a(i_{1}, i_{2}, i_{3}, i_{4}, i_{5}) =$$

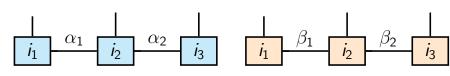
$$\sum g_{1}(i_{1}, \alpha_{1})g_{2}(\alpha_{1}, i_{2}, \alpha_{2})g_{3}(\alpha_{2}, i_{3}, \alpha_{3})g_{4}(\alpha_{3}, i_{4}, \alpha_{4})g_{5}(\alpha_{4}, i_{3})$$

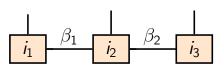
$$= \underbrace{A_{1}^{(i_{1})}}_{1 \times r_{1}} \underbrace{A_{2}^{(i_{2})}}_{r_{1} \times r_{2}} \underbrace{A_{3}^{(i_{3})}}_{r_{2} \times r_{3}} \underbrace{A_{4}^{(i_{4})}}_{r_{3} \times r_{4}} \underbrace{A_{5}^{(i_{5})}}_{r_{4} \times 1}$$



СЛОЖЕНИЕ ТЕНЗОРОВ

$$a(i_1, i_2, i_3) = \underbrace{A_1^{(i_1)}}_{1 \times r_1} \underbrace{A_2^{(i_2)}}_{r_1 \times r_2} \underbrace{A_3^{(i_3)}}_{r_3 \times 1}, \quad b(i_1, i_2, i_3) = \underbrace{B_1^{(i_1)}}_{1 \times s_1} \underbrace{B_2^{(i_2)}}_{s_1 \times s_2} \underbrace{B_3^{(i_3)}}_{s_3 \times 1}$$





$$(a+b)(i_1,i_2,i_3) = \begin{bmatrix} A_1^{(i_1)} & B_1^{(i_1)} \end{bmatrix} \begin{bmatrix} A_2^{(i_2)} & \\ & B_2^{(i_2)} \end{bmatrix} \begin{bmatrix} A_3^{(i_3)} \\ B_3^{(i_3)} \end{bmatrix}$$

ТЕНЗОРНЫЙ ПОЕЗД

специальное представление d-мерной матрицы

$$a(i_1 \dots i_d) = \sum g_1(i_1\alpha_1)g_2(\alpha_1i_2\alpha_2)\dots$$
$$\dots g_{d-1}(\alpha_{d-2}i_{d-1}\alpha_{d-1})g_d(\alpha_{d-1}i_d)$$

d-мерная матрица A представляется через трехмерные матрицы $g_k(\alpha_{k-1}i_k\alpha_k)$ размера $r_{k-1}\times n\times r_k$. Если $r_k\leqslant r$, то число параметров тензорного поезда $\leqslant dnr^2\ll n^d$.

КОГДА ТЕНЗОРНЫЙ ПОЕЗД РЕАЛЬНО ПОМОГАЕТ?

$$A_{k} = [a(i_{1} \dots i_{k}; i_{k+1} \dots i_{d})] =$$

$$\left[\sum u_{k}(i_{1} \dots i_{k}; \alpha_{k}) v_{k}(\alpha_{k}; i_{k+1} \dots i_{d})\right] = U_{k}V_{k}^{\top}$$

$$u_{k}(i_{1} \dots i_{k}\alpha_{k}) = \sum g_{1}(i_{1}\alpha_{1}) \dots g_{k}(\alpha_{k-1}i_{k}\alpha_{k})$$

$$v_{k}(\alpha_{k}i_{k+1} \dots i_{d}) = \sum g_{k+1}(\alpha_{k}i_{k+1}\alpha_{k+1}) \dots g_{d}(\alpha_{k-1}i_{d})$$

Нужно, чтобы все матрицы A_k были близки к матрицам малого ранга.

ТЕНЗОРНЫЙ ПОЕЗД ДЛЯ d-МЕРНОЙ МАТРИЦЫ МОЖНО ПОСТРОИТЬ ПО "МАЛЫМ КРЕСТАМ" В МАТРИЦАХ A_k

$$A(i_1 \dots i_d) = \prod_{k=1}^d A(J_{\leqslant k-1}, i_k, J_{>k}) [A(J_{\leqslant k}, J_{>k})]^{-1}$$

I.Oseledets, E.Tyrtyshnikov'2009

ИСТОРИЯ КРЕСТОВОЙ АППРОКСИМАЦИИ

```
1985 Knuth: Semi-optimal bases for linear dependencies
1995 Tyr., Goreinov, Zamarashkin: A = CGR pseudoskeleton
2000 Tyr.: incomplete cross approximation with ALS maxvol
2000 Bebendorf: ACA = Gaussian elimination
2001 Tyr., Goreinov: maximum volume principle,
     quasioptimality \| \operatorname{cross} \|_{\mathcal{C}} < (r+1) \| \operatorname{best} \|_{2}
2006 Mahoney et al: randomized CUR algorithm
2008 Oseledets, Savostyanov, Tyr.: Cross3D
2009 Oseledets, Tyr.: TT-Cross
2010 J.Schneider: function-related quasioptimality
      \| \operatorname{cross} \|_{C} < (r+1)^{2} \| \operatorname{best} \|_{C}
2011 Tyr., Goreinov: quasioptimality
      \| \operatorname{cross} \|_{C} < (r+1)^{2} \| \operatorname{best} \|_{C}
2013 Ballani, Grasedyck, Kluge: HT-Cross
2013 Townsend, Trefethen -- Chebfun2
```

СТРОЧНО-СТОЛБЦЕВАЯ ИНТЕРПОЛЯЦИЯ МАТРИЦ

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad A_{11} \text{ is } r \times r$$

A интерполируется по первым r строкам и столбцам:

$$A_r = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

ОШИБКА ИНТЕРПОЛЯЦИИ

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} - \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 \\ 0 & A_{22} - A_{21} A_{11}^{-1} A_{12} \end{bmatrix}$$

ПРИНЦИП МАКСИМАЛЬНОГО ОБЪЕМА

ТЕОРЕМА (Горейнов, Тыртышников 2000) Пусть

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \text{ is } r \times r$$

имеет максмимальный объем (определитель по модулю) среди всех $r \times r$ блоков в A, и пусть

$$A_r = \begin{vmatrix} A_{11} \\ A_{21} \end{vmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}.$$

Тогда

$$||A-A_r||_C \leqslant (r+1) \min_{\substack{\mathrm{rank} \ B \leqslant r}} ||A-B||_2.$$

НОВОЕ ПОНЯТИЕ – РЕДУЦИРОВАННЫЙ ОБЪЕМ

$$V_r(A) := \prod_{i=1}^r \sigma_i(A)$$

ТЕОРЕМА (А.И.Осинский – Н.Л.Замарашкин) Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{C}^{M \times N},$$

где $A_{11}\in\mathbb{C}^{m imes n}$ із имеет максимальный r-редуцированный объем среди всех m imes n подматриц и $\mathrm{rank}\,A_{11}\geqslant r$. Тогда

$$||A - CA_{11}^{\dagger}R||_{C} \le \sqrt{1 + \frac{r}{m-r+1}} \sqrt{1 + \frac{r}{n-r+1}} \ \sigma_{r+1},$$

$$C = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \quad R = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}.$$

СЛЕДСТВИЕ ТЕОРЕМЫ О РЕДУЦИРОВАННОМ ОБЪЕМЕ

Взяв крест, в котором на пересечении строк и столюцов находится подматрица размера $(2r-1)\times(2r-1)$, мы можем гарантировать оценку

$$||A-CA_{11}^{\dagger}R||_{\mathcal{C}}\leqslant 2\sigma_{r+1}(A).$$

Достаточно выбрать подматрицу A_{11} с максимальным r-редуцированным объемом.

A.Osinsky, N.Zamarashlin, *Pseudo-skeleton approximations with better accuracy estimates*, submitted to LAA, 2016.



КВАЗИОПТИМАЛЬНОСТЬ ТЕЗОРНОГО ПОЕЗДА, ПОСТРОЕННОГО ПО КРЕСТАМ МАКСИМАЛЬНОГО ОБЪЕМА

$$\widetilde{A}(i_1 \ldots i_d) = \sum_{k=1}^d A(J_{\leqslant k-1}, i_k, J_{>k}) [A(J_{\leqslant k}, J_{>k})]_{\tau_k}^{\dagger}$$

THEOREM.

$$||A - \widetilde{A}||_C \leqslant c(r) \cdot \underbrace{||F||_C}_{\text{BEST APPROXIMATION ERROR}}$$

$$c(r) = \frac{(4r)^{\lceil \log_2 d \rceil} - 1}{4r - 1} (r + 1)^2$$

ТЕНЗОРИЗАЦИЯ ВЕКТОРОВ И МАТРИЦ

Любой вектор размера $N=n_1\dots n_d$ можно рассматривать как d-тензор, а любую $N\times N$ -матрицу

$$a(i,j)=a(i_1\ldots i_d,\ j_1\ldots j_d)$$

можно рассматривать как 2d-тензор или, а еще лучше как d-tensor размера $n_1^2 \times \ldots \times n_d^2$, например такой:

$$a(i_1j_1,\ldots,i_dj_d)$$

Тензоризация плюс тензорный поезд (ТТ) позволяют радикально уменьшить число параметров представления!



БЫСТРОЕ СУММИРОВАНИЕ ЭЛЕМЕНТОВ АСТРОНОМИЧЕСКИ БОЛЬШОГО ВЕКТОРА

$$\sum_{i=1}^{N} a(i) = ? \qquad N = 10^{83}$$

БЫСТРОЕ СУММИРОВАНИЕ ЭЛЕМЕНТОВ АСТРОНОМИЧЕСКИ БОЛЬШОГО ВЕКТОРА

$$i = \overline{i_1 i_2 \dots i_d}$$
 $d = 83$

$$a(i) = a(i_1, \ldots, i_d) = \sum_{\alpha_1, \ldots, \alpha_{d-1}} g_1(i_1, \alpha_1)g_2(\alpha_1, i_2, \alpha_2) \ldots g_d(\alpha_{d-1}, i_d)$$

$$\sum_{i_1,\ldots,i_d} a(i_1,\ldots,i_d) = \sum_{\alpha_1,\ldots,\alpha_{d-1}} \hat{g}_1(\alpha_1)\hat{g}_2(\alpha_1,\alpha_2)\ldots\hat{g}_d(\alpha_{d-1})$$

$$\hat{g}_k = \sum_{i_k} g_k$$



ТТ-ИНТЕГРИРОВАНИЕ

Вычисляется d-мерный интеграл

$$I(d) = \int \sin(x_1 + x_2 + \dots + x_d) \ dx_1 dx_2 \dots dx_d =$$

$$\operatorname{Im} \int_{[0,1]^d} e^{i(x_1 + x_2 + \dots + x_d)} \ dx_1 dx_2 \dots dx_d = \operatorname{Im} \left(\left(\frac{e^i - 1}{i} \right)^d \right).$$

n=11 узлов по каждому измерению \Rightarrow всего n^d значений! Для построения ТТ-интерполяции из них реально вычисляется лишь малая часть.

d	I(d)	Относит. погрешность	Время
1000	-2.637513e-19	3.482065e-11	11.60
2000	2.628834e-37	8.905594e-12	33.05
4000	9.400335e-74	2.284085e-10	105.49

ТЕНЗОРИЗАЦИЯ ПРИ ВЫЧИСЛЕНИИ ОДНОМЕРНЫХ ИНТЕГРАЛОВ

Интеграл

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$$

приближается интегралом по конечному отрезку, а последний вычисляется по формуле прямоугольников.

Для машинной точности нужно порядка 2^{77} значений. Вектор значений рассматривается как тензор размеров $2\times 2\times \ldots \times 2$.

При ТТ-интерполяции с машинной точностью ТТ-ранги $\leqslant 12$. Меньше 1 сек. на ноутбуке.

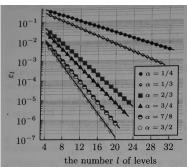


TT-FE APPROXIMATION

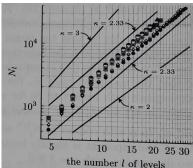
$$u_{\Gamma}(x) = r^{\alpha} \sin \alpha \phi(x), \quad x \in \Omega = (0, 1)^2$$

 $\varepsilon_I \leqslant \exp\{-cN_I^{\frac{1}{\kappa}}\}, \quad N_I$ – the number of TT-elements

THEOREM (V.Kazeev & C.Schwab). $\kappa \leqslant 5$.



(a) Convergence with respect to l. The reference lines correspond to the exponential convergence $\varepsilon_l = C_\alpha \, 2^{-\tilde{\alpha} \, l}$ with C_α independent of l and with $\tilde{\alpha} = \min\{\alpha, 1\}$.



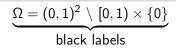
(d) The number N_l (3.3.2) of QTT parameters vs. l. The reference lines correspond to the algebraic growth $N_l = C_{\alpha} \, l^{\kappa}$ with κ and C_{α} independent of l.

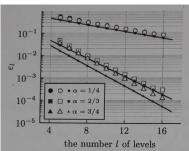


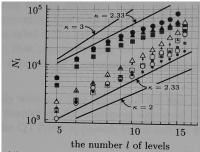
TE-FE-AMEN

$$\Delta u = 0$$
, $u|_{\partial\Omega} = u_{\Gamma}$

$$\underbrace{\Omega = (0,1)^2 \, \setminus \, [0,1) \times (-1,0]}_{\text{bright labels}}$$







(a) Convergence with respect to l. The reference lines correspond to the exponential convergence $\varepsilon_l = C_\alpha \, 2^{-\alpha l}$ with C_α independent of l. The markers for $u_{\rm sol}^l$ and $u_{\rm tr}^l$ mostly coincide,

(d) The number N_l (3.3.2) of QTT parameters vs. l. The reference lines correspond to the algebraic growth $N_l = C_{\alpha} l^{\kappa}$ with κ and C_{α} independent of l.

(V.Kazeev)

LOW RANK AND TENSOR TRAINS IN THE SMOLUCHOWSKI EQUATIONS

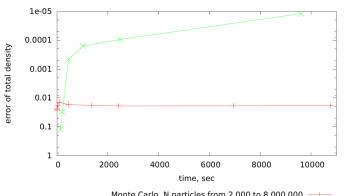
- $\overline{v} = (v_1, \dots, v_d)$ volumes of different substances of a particle
- ▶ t time
- ▶ $n(\overline{v}, t)$ concentration function for volume components of a particle

$$\frac{\partial n(\overline{v},t)}{\partial t} = \frac{1}{2} \int_0^{v_1} du_1 \dots \int_0^{v_d} K(\overline{v} - \overline{u}; \overline{u}) n(\overline{u},t) n(\overline{v} - \overline{u},t) du_d - \\
-n(\overline{v},t) \int_0^{\infty} du_1 \dots \int_0^{\infty} K(\overline{v}; \overline{u}) n(\overline{u},t) du_d, \\
n(\overline{v},0) = n_0(\overline{v}).$$

Joint work with S.Matveev and A.Smirnov



TENSOR TRAIN VS MONTE CARLO FOR 2D SMOLUCHOWSKI EQUATIONS



Ballistic kernel:
$$K(u, v) = (u^{\frac{1}{3}} + v^{\frac{1}{3}})^2 \sqrt{\frac{1}{u} + \frac{1}{v}}$$
.



ТЕНЗОРНЫЙ ПОЕЗД КАК ИНСТРУМЕНТ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ

TEOPEMA. Если A_{\blacksquare} имеет максимальный объем среди всех $r \times r$ блоков в A, то

$$||A_{\blacksquare}||_C \ge ||A||_C/(2r^2+r).$$



S. Goreinov, I. Oseledets, D. Savostyanov, E. Tyrtyshnikov,

N. Zamarashkin, How to find a good submatrix, Matrix Methods: Theory,

Algorithms and Applications. Devoted to the Memory of Gene Golub (eds.

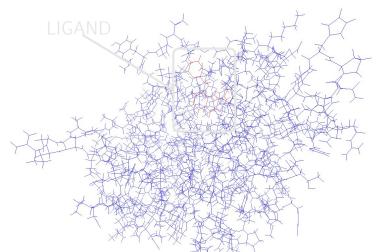
V.Olshevsky and E.Tyrtyshnikov), World Scientific Publishers, Singapore,

2010, pp. 247–256.



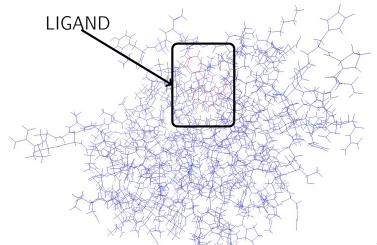
ПРЯМОЙ ДОКИНГ ПРИ РАЗРАБОТКЕ ЛЕКАРСТВ

ACCOMMODATION OF LIGAND INTO PROTEIN

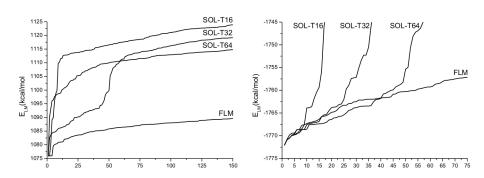


ПРЯМОЙ ДОКИНГ ПРИ РАЗРАБОТКЕ ЛЕКАРСТВ

ACCOMMODATION OF LIGAND INTO PROTEIN



DIRECT DOCKING IN THE DRUG DESIGN



Joint work with D.Zheltkov and V.Sulimov

WELCOME THE BLESSING OF DINENSIONALITY

- Fokker-Planck, Smoluchovski equations
- Differential equations with parameters
- Green functions in integral equations
- Spin dynamics
- Global optimization algorithms
- Many others

RECENT BOOKS:

- G. Golub and Ch. Van Loan, Matrix Computations, 4th edition, 2013.
- W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus, Springer, 2012.



