Вычислительные технологии решения задач оптимизации с разреженной структурой

А.С. Аникин
anikin@icc.ru

Институт динамики систем и теории управления
имени В.М. Матросова СО РАН, Иркутск

27.11.2019

$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$\|x^{k+1} - x^k\|_0 = n, \qquad O(n)$$

$$\downarrow$$

$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$\|x^{k+1} - x^k\|_0 = n, \qquad O(n)$$

$$\downarrow$$

$$\|x^{k+1} - x^k\|_0 = s \ll n, \qquad O(s)$$

$$\downarrow$$

$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$\|x^{k+1} - x^k\|_0 = n, \qquad O(n)$$

$$\downarrow$$

$$\|x^{k+1} - x^k\|_0 = s \ll n, \qquad O(s)$$

$$\downarrow$$

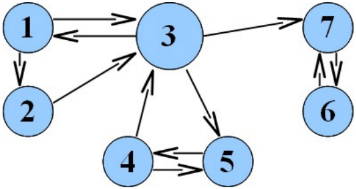
$$\|x^{k+1} - x^k\|_0 = 1, \qquad O(1)$$

- Разреженность решения
- Очень "дешевые" итерации

Задача PageRank

$$P^T x = x$$
$$P \in \mathbb{R}^{n \times n}, \;\; x \in \mathbb{R}^n_+$$
$$\langle x, e \rangle = 1, \;\; e = (1, ..., 1)^T$$

где P - стохастическая матрица



$$P^T = \begin{pmatrix} 0 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/3 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$f(x) = \frac{1}{2}||Ax||_2^2 \to \min_{\langle x,e\rangle=1}$$

$$f(x) = \frac{1}{2}||Ax||_2^2 + \frac{\gamma}{2}(\langle x, e\rangle - 1)^2 \to \min$$

где:

$A = P^T - I$, I - единичная матрица;

γ - параметр штрафа за нарушение ограничения $\langle x, e\rangle = 1$.

- "Полные" (Классические) методы: FGM, CG, BB, Polyak.
- "Разреженные" методы: RCD, FW, NL1.

Для ускорения классических градиентных методов применяется GPU.

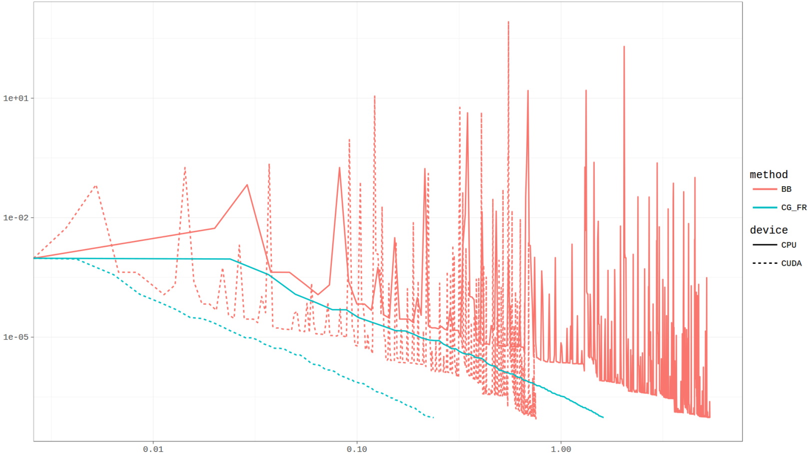
Характеристики матрицы A, построенной по базам Стэнфордского университета

web-graph		Число ненулевых элементов				
		в строке		в столбце		среднее
		мин.	макс.	мин.	макс.	
Stanford,	$n = 281903$	2	38607	1	256	9.2
NotreDame,	$n = 325729$	2	10722	1	3445	5.51
BerkStan,	$n = 685230$	1	84209	1	250	12.09
Google,	$n = 875713$	1	6327	1	457	6.83

Время минимизации, сек.

web-graph	CG			BB		
	CPU	GPU	$\frac{CPU}{GPU}$	CPU	GPU	$\frac{CPU}{GPU}$
Stanford	1.61	0.14	*11.50*	5.39	0.50	*10.78*
NotreDame	27.78	2.76	*10.05*	61.81	7.12	*8.68*
BerkStan	5.49	0.60	*9.15*	18.22	2.57	*7.08*
Google	52.47	2.97	*17.66*	176.91	5.84	*30.29*

Задача Stanford, сходимость методов



FW — Метод условного градинта (Frank-Wolfe)

$$x_{k+1} = (1 - \gamma_k)x_k + \gamma_k y_k, \ \gamma_k = \frac{2}{k+1}, \ k = 1, 2, ...$$

$$\langle \nabla f(x_k), y \rangle \to \min_{y \in S_n(1)}$$

$$y_k = (0, ..., 0, 1, 0, ..., 0),$$

где 1 находится в позиции

$$i_k = \operatorname*{argmin}_{i=1,...,n} \partial f(x_k)/\partial x^i$$

Итерация выполняет 1 обновление значения функциии и градиента.

NL1 — Градиентный метод в L1-норме

$$x_{k+1} = x_k + h \cdot y_k$$

$$h = \frac{1}{L}(g_{\max} - g_{\min}) = \frac{1}{3}(g_{\max} - g_{\min})$$

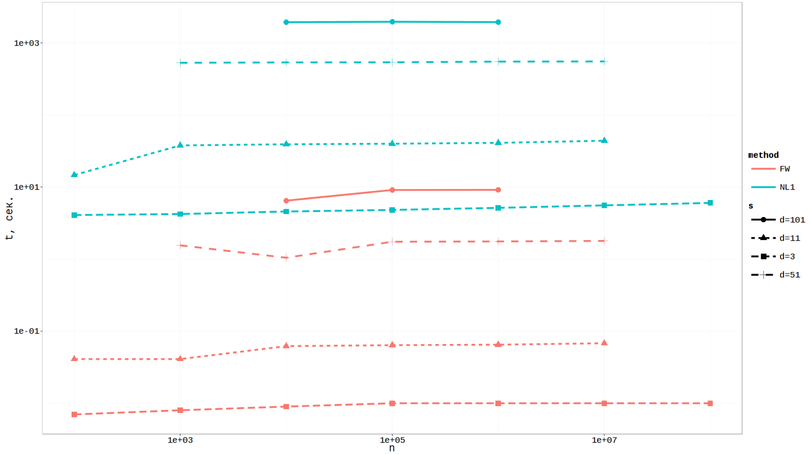
$$y = (0, ..., 0, 1^{\max}, 0, ...0, 1^{\min}, 0, ...0), ||y||_0 = 2$$

$$g_{\max} = \underset{i=1,\dots,n}{\operatorname{argmax}} \; \partial f(x_k)/\partial x^i$$

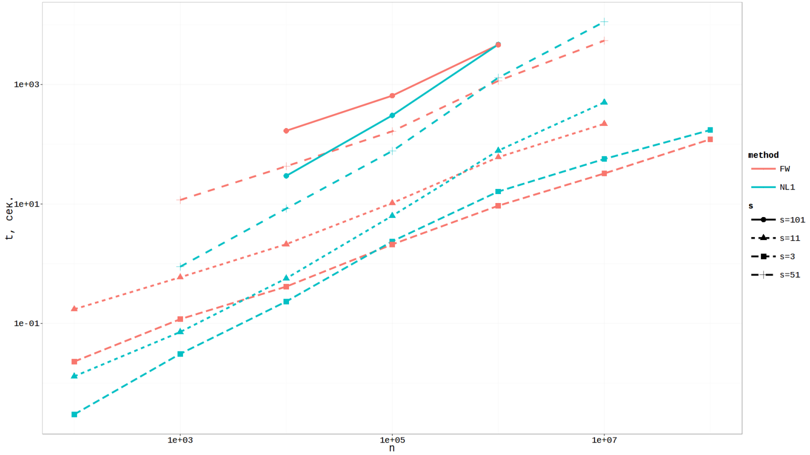
$$g_{\min} = \underset{i=1,\dots,n}{\operatorname{argmin}} \; \partial f(x_k)/\partial x^i$$

Итерация выполняет 2 обновления значения функциии и градиента.

Диагональные матрицы



Матрицы со случайной структурой



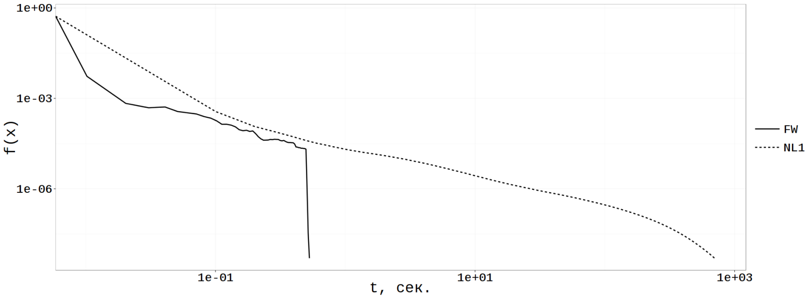
Время минимизации, сек.

graph	n	NL1		FW	
		t	iterations	t	iterations
Stanford	281903	0.145	93152	0.008	14142
NotreDame	325729	700.810	3816436	0.526	38014
BerkStan	685230	38161.847	12315700	0.536	19990
Google	875713	113.643	1083996	0.278	37313

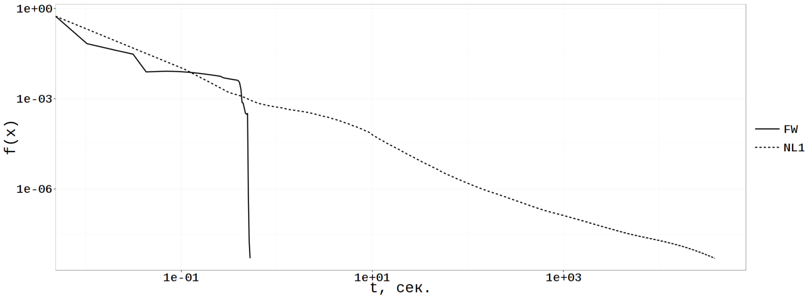
"Стоимость" итерации

		Stanford		BerkStan	
		NL1	FW	NL1	FW
s_r	min	1.0	1.0	1.0	1.0
	max	34.0	4.0	84209.0	84209.0
	med	3.9	3.9	2278.4	148.6
s_c	min	2.0	2.0	1.0	1.0
	max	37.0	3.0	244.0	83.0
	med	2.9	2.8	15.7	6.2
$s_r \cdot s_c$	min	3.0	3.0	2.0	2.0
	max	1258.0	12.0	15494456.0	6989347.0
	med	11.7	11.3	84304.3	7507.5

web-NotreDame



BerkStan



Задача восстановления матрицы корреспонденций компьютерных сетей

$$f(x) = \|x - x_g\|_2^2 \to \min_{Ax=b},$$

x_g – гравитационная точка,

$$A \in \mathbb{R}^{m \times n}, \quad A_{ij} \in \{0, 1\}.$$

$$f(x) = \|x\|_2^2 \to \min_{Ax=b}.$$

$$f(x) = \frac{1}{2}\|x\|_2^2 + \frac{\gamma}{2}\|Ax - b\|_2^2 \to \min_x,$$

γ – штрафной параметр.

Найденное решение x считается корректным, если специальные вспомогательные функции $QLLA(x)$ и $sumQLLA(x)$ («метрики качества») имеют значение, превышающее некоторый заданный порог.

Для получения решения с требуемыми свойствами применена вычислительная схема, комбинирующая работу параллельной версии метода сопряжённых градиентов (CG_FR) и покомпонентного варианта метода покоординатного спуска (CDp).

5000 nodes, $n = 24995000 \approx 25 \cdot 10^6$, загрузка данных $\approx$ 90 сек.
Требуемая память:

$A =$ 8.266 Gb,
$A^T =$ 8.452 Gb,
$b =$ 313.875 Kb,
$x =$ 190.697 Mb.

Метод	t, сек.	QLLA	sumQLLA
условие остановки: (sumQLLA $\geq 90\%$)			
CG_FR	30.060	7.091	23.981
CDp	4.201	34.550	90.082
условие остановки: (QLLA $\geq 90\%$)			
CG_FR	29.615	7.091	23.981
CD	157.323	90.014	99.906

10000 nodes, $n = 99990000 \approx 1 \cdot 10^8$, загрузка данных $\approx$ 490 сек.

Требуемая память:

$A =$ 34.688 Gb,

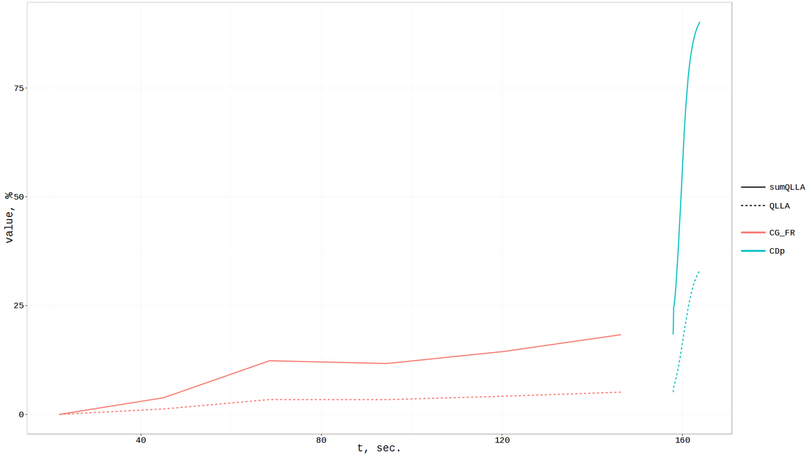
$A^T =$ 35.432 Gb,

$b =$ 627.938 Kb,

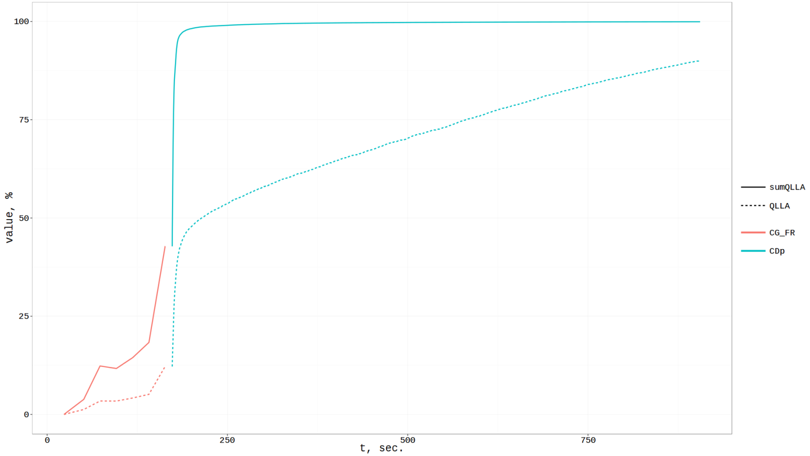
$x =$ 762.863 Mb.

Метод	t, сек.	QLLA	sumQLLA
условие остановки: (sumQLLA $\geq 90\%$)			
CG_FR	146.358	5.116	18.317
CD	17.504	33.184	90.224
условие остановки: (QLLA $\geq 90\%$)			
CG_FR	163.634	12.198	42.785
CD	741.753	90.002	99.945

10000 nodes, sumQLLA $\geq 90\%$



10000 nodes, QLLA $\geq 90\%$



Ускоренный покомпонентный метод

$$\frac{1}{2}\|Ax\|_2^2 \to \min_{\substack{x\in\mathbb{R}^n_+,\\ \langle x,e\rangle=1}}$$

$$\|x-\tilde{x}\|_2 \to \min_{\substack{x\in\mathbb{R}^n,\\ Ax=b}}$$

$$\ln\left(\sum_{k=1}^{m}\exp\left(\langle a_k, x\rangle\right)\right) \to \min_{x\in\mathbb{R}^n}$$

$$\|a_k\|_0 = s \ll n$$

Известные ускоренные покомпонентные методы – стоимость итерации порядка $O(n)$.

Наша задача – стоимость итерации порядка $O(s)$.

За основу берём рандомизированный покомпонентный (координатный) метод Ю.Е. Нестерова:

$$x^{k+1} = x^k - \frac{1}{L_{i^k}} \nabla_{i^k} f(x^k)$$

$$p(i^k) = \frac{L_i}{\sum_{j=1}^n L_j}$$

$$|\nabla_i f(x + \delta e_i) - \nabla_i f(x)| \leq L_i |\delta|$$

Вполне очевидно, что амортизационная сложность итерации неускоренного метода Нестерова для рассматриваемого класса задач будет равна $O(s)$ – это стоимость пересчета случайно выбранной компоненты градиента.

Для ускорения применяем технику "Каталист":

$$F_{x^k,\bar{L}}(y) = f(y) + \frac{\bar{L}}{2}\|y - x^k\|_2^2 \to \min_y$$

$$f_L(x^k) = \min_y F_{x^k,\bar{L}}(y) = F_{x^k,\bar{L}}(y_{\bar{L}}(x^k))$$

$$\nabla f_{\bar{L}}(x^k) = \bar{L}(x^k - y_{\bar{L}}(x^k))$$

$$O(\sqrt{\frac{\bar{L}R^2}{\varepsilon}})$$

Чем меньше выбирается параметр $\bar{L}$, тем эта оценка будет лучше, но при этом тем сложнее на каждой итерации решать вспомогательную задачу.

Общая трудоемкость предлагаемого ускоренного метода будет равна:

$$N_C \cdot (C_g + N_N \cdot C_N),$$

где:

N_C – число внешних итераций "Каталиста",

C_g – стоимость расчета полного градиента оптимизируемого функционала,

N_N – число итераций неускоренного метода Нестерова для решения вспомогательной задачи на каждой внешней итерации «Каталиста»,

C_N – амортизационная стоимость итерации неускоренного метода.

При правильном выборе регуляризующего параметра

$$\bar{L} = \frac{1}{n}\sum_{i=1}^{n} L_i$$

для рассматриваемой задачи имеем:

$$N_C = O(\sqrt{\bar{L}R^2/\varepsilon}),$$

$$C_g = O(sn),$$

$$N_N = \tilde{O}(n),$$

$$C_N = O(s).$$

Резюмируя всё вышесказанное, получаем, что для предлагаемого метода общее число итераций будет равно:

$$\tilde{O}\left(\sqrt{n\sum_{i=1}^{n} L_i R^2/\varepsilon}\right)$$

Если бы был использован ускоренный покомпонентный метод то для него мы бы получили лучшую оценку на число итераций:

$$O\left(\sum_{i=1}^{n}\sqrt{L_i}\sqrt{R^2/\varepsilon}\right)$$

$$\sum_{i=1}^{n}\sqrt{L_i} \leq \sqrt{n\sum_{i=1}^{n} L_i}$$

Предлагаемый метод должен быть эффективен для сильно разреженных постановок, когда потери из-за возросшего числа итераций с запасом компенсируются их низкой стоимостью.

Вычислительные технологии решения задач оптимизации с разреженной структурой

А.С. Аникин
anikin@icc.ru

Институт динамики систем и теории управления
имени В.М. Матросова СО РАН, Иркутск

27.11.2019