

О СПОСОБАХ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ СВЕРХБОЛЬШИХ РАЗМЕРОВ

Извлекаем пользу из Big Data



РУКОВОДИТЕЛЬ:
Александр Гасников

УЧАСТНИКИ:
Даниил Смирнов
Сергей Ким
Степан Плаунов
Федор Носков

ПРОБЛЕМА

Быстрое решение системы линейных уравнений с n неизвестными, $n \gg 1$

АКТУАЛЬНОСТЬ

PageRank
Truss topology design
Compressed sensing

СУЩЕСТВУЮЩИЕ РЕШЕНИЯ

Метод простых итераций
Методы Монте-Карло
Градиентный спуск
Сопряженные градиенты
LBFGS

ВОЗМОЖНЫЕ РЕШЕНИЯ

Быстрые градиентный, покомпонентный и двойственный спуски с адаптивным подбором шага
Модификация метода условного градиента

ПЛАН РАБОТЫ

- Изучение марковских процессов, вывод PageRank
- Изучение современных численных методов выпуклой оптимизации
- Практическая реализация
- Экспериментальное обоснование степенного закона убывания компонент вектора PageRank

МАТЕРИАЛЫ И ОБОРУДОВАНИЕ

Ноутбуки, съемные носители, библиотеки numpy и scipy для python 3.5

Lenovo

B50-45

Имя компьютера

LAPTOP-CH28GBBF

Переименование компьютера

Организация

WORKGROUP

Выпуск

Windows 10 Домашняя для одного языка

Код продукта

00327-30345-93159-AAOEM

Серийный номер

CB36514323

Процессор

AMD A6-6310 APU with AMD Radeon R4
Graphics 1.80 GHz

Установленная ОЗУ

6,00 ГБ (доступно: 4,96 ГБ)

Тип системы

64-разрядная операционная система,
процессор x64

Перо и сенсорный ввод

Для этого монитора недоступен ввод с
помощью пера и сенсорный ввод

ОБЛАСТЬ ПРИМЕНЕНИЯ PAGERANK

сириус



Найти

Войти

Сириус

sochisirius.ru ▾

День России в Образовательном центре «Сириус». 27. 05 искусство. Владимир Кудря: «Конечный результат – красота».

Условия размещения Критерии отбора Все фото Новости Наука

8 (800) 100-76-63 · ежедневно, круглосуточно

📍 Краснодарский край, Сочи г., Олимпийский просп., 40

Сириус — Википедия

ru.wikipedia.org > Сириус ▾

Си́риус (лат. **Sirius**), также α Большого Пса (лат. α Canis Majoris) — ярчайшая звезда ночного неба. Хотя его светимость и превышает в 22 раза светимость Солнца, она отнюдь не является рекордной в мире звёзд — высокий видимый блеск **Сириуса**...

Сириус — НОВОСТИ



Владимир Путин: аналоги центра «**Сириус**» до...

kuban.aif.ru 11:46 вчера

Президент России 19 июля побывал с визитом в уникальном образовательном центре в Сочи

Путин улетел на «**Сириус**» и встретился там...

mk.ru 19 июл 2016

Президент оценил работу «**Сириуса**»

krasnodar.bz 4:04 вчера



Сириус

Звезда

Ярчайшая звезда ночного неба. Хотя его светимость и превышает в 22 раза светимость Солнца, она отнюдь не является рекордной в мире звёзд - высокий видимый блеск Сириуса обусловлен его относительной близостью. Сириус можно наблюдать из любого региона... [Читать дальше](#)

Масса: 3.98×10^{30} кг

Смотрите также



Бетельгейзе



Альфа
Центавра



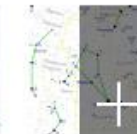
Большой
Пёс



Альдебаран



Процион



Арктур

PAGERANK

1998 – предложена модель ранжирования веб-страниц согласно вектору PageRank (Ларри Пейдж, Сергей Брин)

$$\pi^T = \pi^T P$$

n – ЧИСЛО ВЕРШИН (WEB-СТРАНИЦ)

M – ЧИСЛО ЧЕЛОВЕЧКОВ ($M \geq 1$)

$\forall t \geq \frac{1}{\alpha} \ln \frac{n}{\epsilon}$
spectral gap P

$\pi: \pi^T P = \pi^T$
 $\|P_{ij}\|_{j=1}^n$

$P \left(\left\| \frac{n(t)}{M} - \pi \right\|_2 \geq \frac{2\sqrt{2} + 4\sqrt{\ln 5^{-1}}}{\sqrt{M}} \right) \leq \epsilon$

$n(t) = (n_1(t), \dots, n_n(t))^T$; $n_k(t)$ – ЧИСЛО ЧЕЛОВЕЧКОВ В МОМЕНТ ВРЕМЕНИ t В ВЕРШИНЕ k .

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

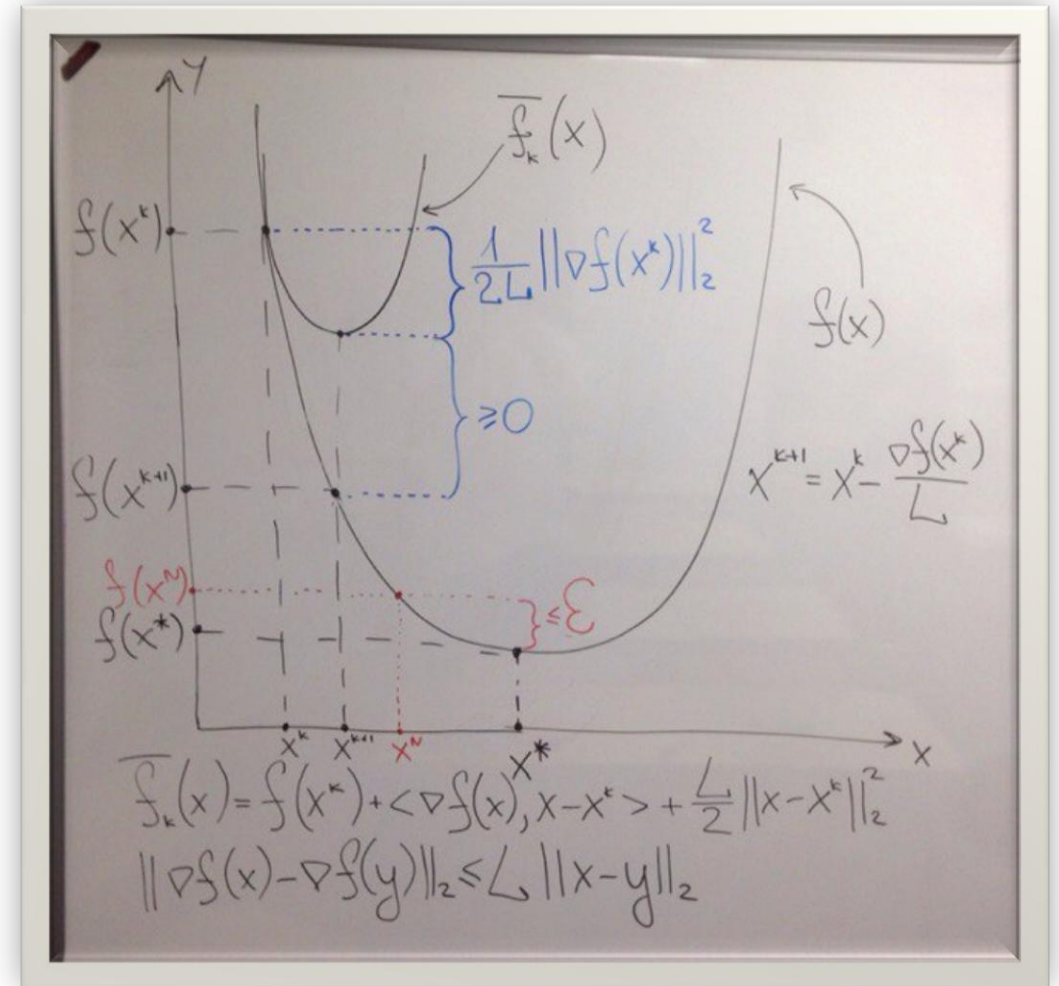
(I) $Ax = b$
 $(P^T - I)x = b$
 $x^T(P - I) = b^T$
 $x^T P = x^T + b^T$

(II) $Ax - b = 0$
 $\frac{1}{2} \|Ax - b\|_2^2 = 0$
 $f(x) = 0$

$A = \begin{pmatrix} P^T - I \\ \Delta & \dots & \Delta \end{pmatrix}$
 $P = \|p_{ij}\|_{i,j=1, \dots, n}^{n, n}$
 $i \circ \xrightarrow{p_{ij}} \circ j$
 $b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \Delta \end{pmatrix}$
 $I = \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix}$

Свели к задаче
 $f(x) \rightarrow \min$

PageRank и система линейных уравнений



Визуализация метода градиентного спуска

ВРЕМЯ РАБОТЫ МЕТОДОВ

ϵ – точность ($= (\|Ax^N - b\|_2)^2$)

n – количество веб-страниц

α – spectral gap P ($\alpha \ll 1$)

$R^2 = (\|x^* - x^0\|_2)^2$

s – среднее число ненулевых элементов в столбце

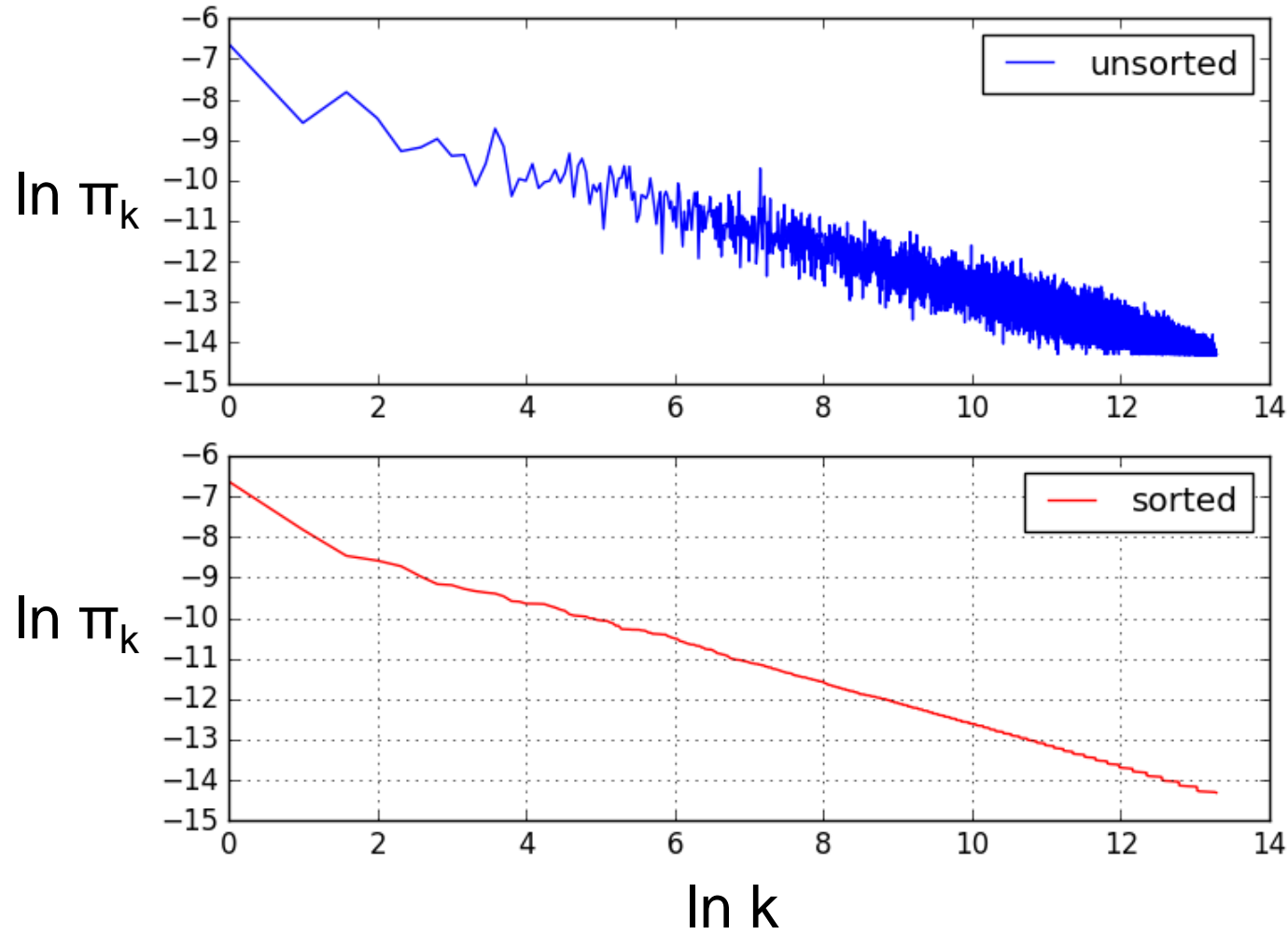


Встроенные в Python 3.5

Методы	Количество итераций N	Стоимость итерации	Время работы (секунды)
Power Method	$\frac{1}{\alpha} \ln \frac{1}{\epsilon}$	sn	0.93
Random Walk	$\frac{1}{\alpha} \ln \frac{n}{\epsilon}$	$\frac{1}{\epsilon^2} \ln n$	1.38
Gradient Method (GM)	LR^2 / ϵ	sn	1.51
Fast GM (FGM)	$\sqrt{LR^2 / \epsilon}$	sn	0.82
Coordinate GM	nLR^2 / ϵ	n	>100
Coordinate FGM	$n\sqrt{LR^2 / \epsilon}$	n	8.97
Conditional GM	LR^2 / ϵ	sn	>100
Dual FGM	?	sn	8.63
Conjugated GM	$\sqrt{LR^2 / \epsilon}$	sn	1.15
LBFGS	?	$\gg sn$	>100

Web-граф получен по модели Боллобаша–Риордана $\epsilon = 10^{-5}, n = 10^4$

ВЫВОДЫ И АНАЛИЗ РЕЗУЛЬТАТА



Авраченков–Лебедев 2006

Модель Боллобаша–Риордана =
модель Бакли–Остгуса ($a = 1$)

$\pi = [\pi_1, \dots, \pi_n]^T$ – PageRank

$$E[\pi_k] \sim k^{-\gamma}$$

$$\gamma \approx 0.5$$

ВЫВОДЫ И АНАЛИЗ РЕЗУЛЬТАТА

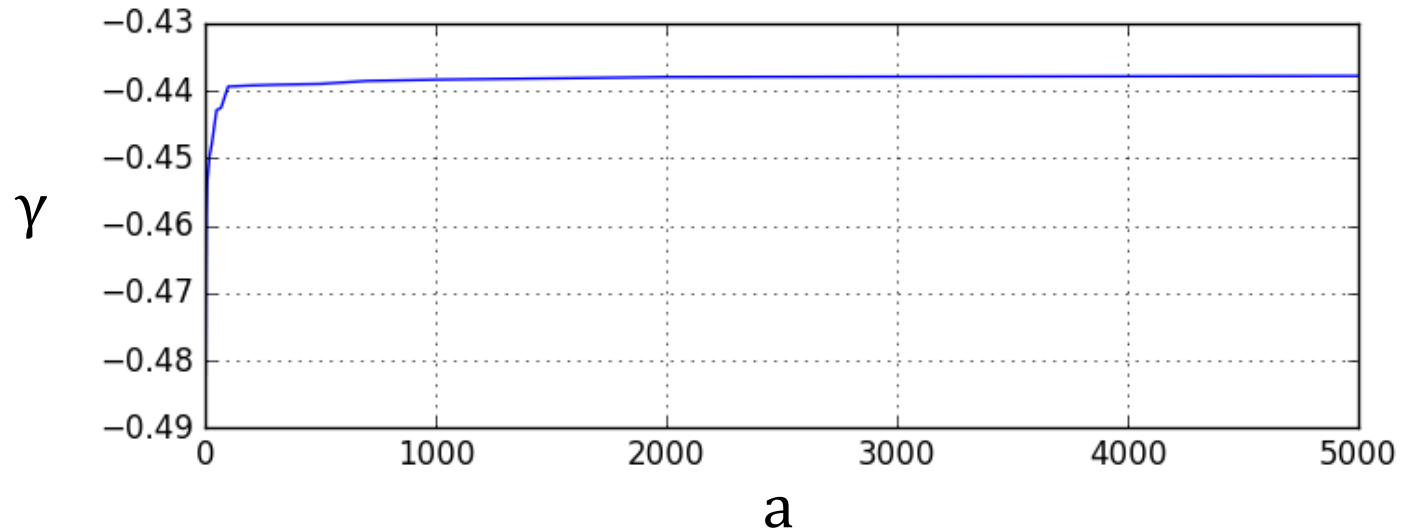
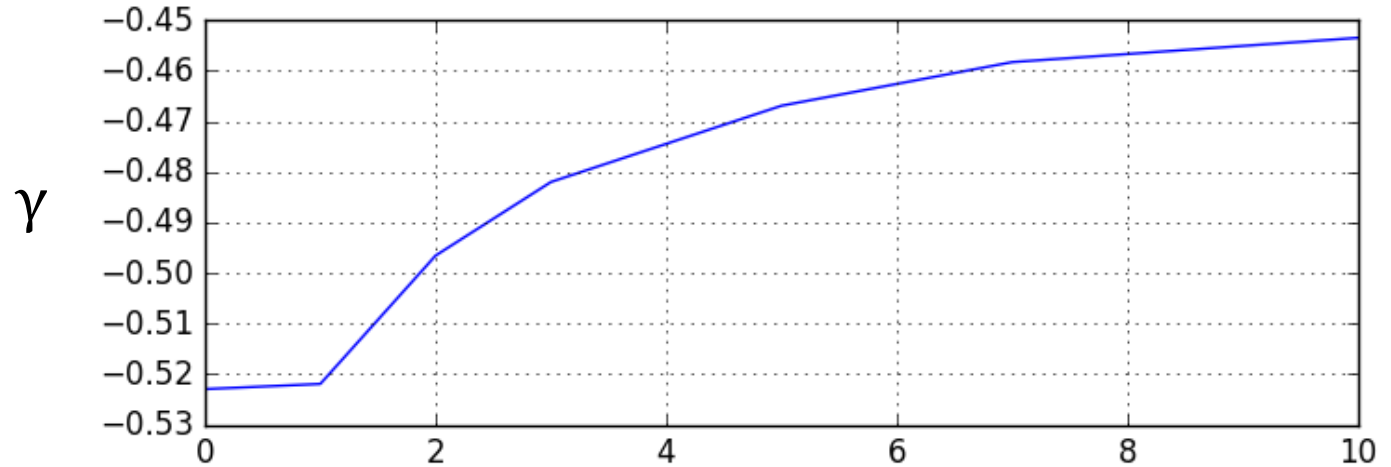
График зависимости показателя степени в степенном законе убывания компонент вектора PageRank от параметра a модели Бакли–Остгуса

$$\pi_k \sim k^{-\gamma(a)}$$

$$\gamma(0) \approx 0.52$$

$$\gamma(1) \approx 0.50$$

$$\gamma(5000) \approx 0.44$$



ИТОГОВЫЙ РЕЗУЛЬТАТ

- База собранных данных
- Набор реализованных алгоритмов
- Исследование закона распределения степеней вершин PageRank



Фотография проектной команды

СПАСИБО ЗА ВНИМАНИЕ

<https://github.com/Kolldun/PageRank-gradient>