A problem generator for big data optimization

Kimon Fountoulakis and Jacek Gondzio

Optimization and Big Data 2015

Motivation - need for controlled testing

Big data/large-scale sparked the development of new optimization methods

Percentage of big data/large-scale papers within convex optimization since 2012

- ArXiv: 53%
- Optimization Online: 43%
- Journal of Machine Learning Research: 60%
- Springer: 51%

Issue: frequently, the performance of new methods is tested on well-conditioned randomly generated problems

Need: controlled testing – a problem generator which can reveal weaknesses and strengths of new methods

Contribution

A problem generator for

$$\text{minimize } \tau\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

$\tau > 0$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

Contribution

A problem generator for

$$\text{minimize } \tau\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

$\tau > 0$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

Inexpensive

Contribution

A problem generator for

$$\text{minimize } \tau\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

$\tau > 0$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

Low memory requirements

Contribution

A problem generator for

$$\text{minimize } \tau\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

$\tau > 0$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

n	processors	terabytes	seconds
$2^{36} \approx 68$ billion	4096	12.288	1970
2^{38}	16384	49.152	1990
$2^{40} \approx 1$ trillion	65536	196.608	2006

Flexibility

The generator allows control of the

Flexibility

The generator allows control of the

- dimensions m, n

Flexibility

The generator allows control of the

- dimensions m, n
- sparsity of A

Flexibility

The generator allows control of the

- dimensions m, n
- sparsity of A
- sparsity of $A^\intercal A$ (independently of A)

Flexibility

The generator allows control of the

- dimensions m, n
- sparsity of A
- sparsity of $A^\intercal A$ (independently of A)
- singular value decomposition of A

Flexibility

The generator allows control of the

- dimensions m, n
- sparsity of A
- sparsity of $A^\intercal A$ (independently of A)
- singular value decomposition of A
- sparsity of the optimal solution x^*

Flexibility

The generator allows control of the

- dimensions m, n
- sparsity of A
- sparsity of $A^\intercal A$ (independently of A)
- singular value decomposition of A
- sparsity of the optimal solution x^*
- components of x^*

Details are presented in the poster

Software

- MATLAB implementation: up to 4 million variables on a PC with 8 GB RAM
- MPI (Message Passing Interface) implementation for distributed computing

Search on Google: ERGO trillion