Dantzig's Simplex Algorithm

How to Write Fast Numerical Code

by Rico Häuselmann

Donjan Rodic

Swiss Federal Institute of Technology (ETH Zurich)

27.05.2013

Linear Programming

Optimising a Linear Program in standard form:

```
Maximize
2x - 3y + z

Subject To
 x  +  y + z <= 10
4x  - 3y + z <=  3
2x  +  y - z <= 6
```

Restrictions

- all coefficients positive (simplicity)
- all coefficients $\leq 10^6$ (stability)

Steps

- Tableau form

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 4 & -3 & 1 & 0 & 1 & 0 & 0 & 3 \\ 2 & 1 & -1 & 0 & 0 & 1 & 0 & 6 \\ 2 & -3 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- Pivoting, reduced cost (objective function)
- Termination, worst runtime $O(e^m)$, but often $O(m)$

Steps

- Tableau form $\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 4 & -3 & 1 & 0 & 1 & 0 & 0 & 3 \\ 2 & 1 & -1 & 0 & 0 & 1 & 0 & 6 \\ 2 & -3 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$
- Pivoting, reduced cost (objective function)
- Termination, worst runtime $O(e^m)$, but often $O(m)$

Steps

- Tableau form $\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 4 & -3 & 1 & 0 & 1 & 0 & 0 & 3 \\ 2 & 1 & -1 & 0 & 0 & 1 & 0 & 6 \\ 2 & -3 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$
- Pivoting, reduced cost (objective function)
- Termination, worst runtime $O(e^m)$, but often $O(m)$

Comparison

- GLPK (GNU Linear Programming Kit), solid standard solver
- CPPLEX, mathematical OO-implementation
- Gurobi (CPLEX), fastest (multithreaded) solver available
- SoPlex, fastest FOSS solver available

Comparison

- GLPK (GNU Linear Programming Kit), solid standard solver
- CPPLEX, mathematical OO-implementation
- Gurobi (CPLEX), fastest (multithreaded) solver available
- SoPlex, fastest FOSS solver available

Comparison

- GLPK (GNU Linear Programming Kit), solid standard solver
- CPPLEX, mathematical OO-implementation
- Gurobi (CPLEX), fastest (multithreaded) solver available
- SoPlex, fastest FOSS solver available

Comparison

- GLPK (GNU Linear Programming Kit), solid standard solver
- CPPLEX, mathematical OO-implementation
- Gurobi (CPLEX), fastest (multithreaded) solver available
- SoPlex, fastest FOSS solver available

Properties

- Tableau: $(m+1) \times (m+n+2)$
  (requires full access each iteration)
  Memory reads: $m(m+n)+2m+n$
  (all capacity misses for bigger problems)
  Flops: $2 * m(m+n)+m$
- Computational intensity $I = \frac{2m^2+2mn+4m+2n}{8(m^2+mn)} \sim \frac{1}{4}$

Properties

- Tableau: $(m+1) \times (m+n+2)$
  (requires full access each iteration)
  Memory reads: $m(m+n) + 2m + n$
  (all capacity misses for bigger problems)
  Flops: $2 * m(m+n) + m$
- Computational intensity $I = \frac{2m^2+2mn+4m+2n}{8(m^2+mn)} \sim \frac{1}{4}$

Implementation

- baseline
- nta: cache control (inhibit polluting tableau data, prefetch next row)
- ssa: static scalar assignment
- block: reuse pivot row
- block_swap: swap pivot row at the end of tableau

Implementation

- baseline
- nta: cache control (inhibit polluting tableau data, prefetch next row)
- ssa: static scalar assignment
- block: reuse pivot row
- block_swap: swap pivot row at the end of tableau

Implementation

- baseline
- nta: cache control (inhibit polluting tableau data, prefetch next row)
- ssa: static scalar assignment
- block: reuse pivot row
- block_swap: swap pivot row at the end of tableau

Implementation

- baseline
- nta: cache control (inhibit polluting tableau data, prefetch next row)
- ssa: static scalar assignment
- block: reuse pivot row
- block_swap: swap pivot row at the end of tableau

Implementation

- baseline
- nta: cache control (inhibit polluting tableau data, prefetch next row)
- ssa: static scalar assignment
- block: reuse pivot row
- block_swap: swap pivot row at the end of tableau

Performance

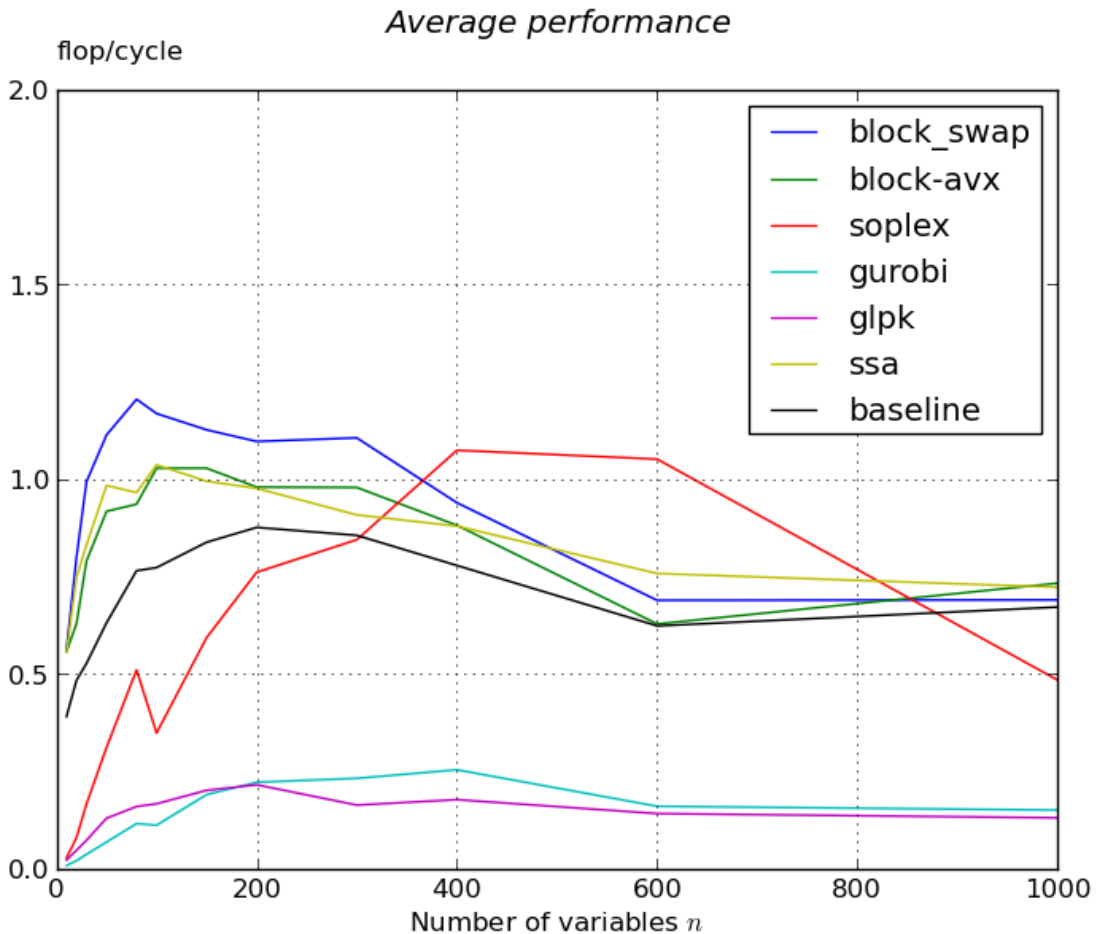


Figure : performance comparison

Wall Time

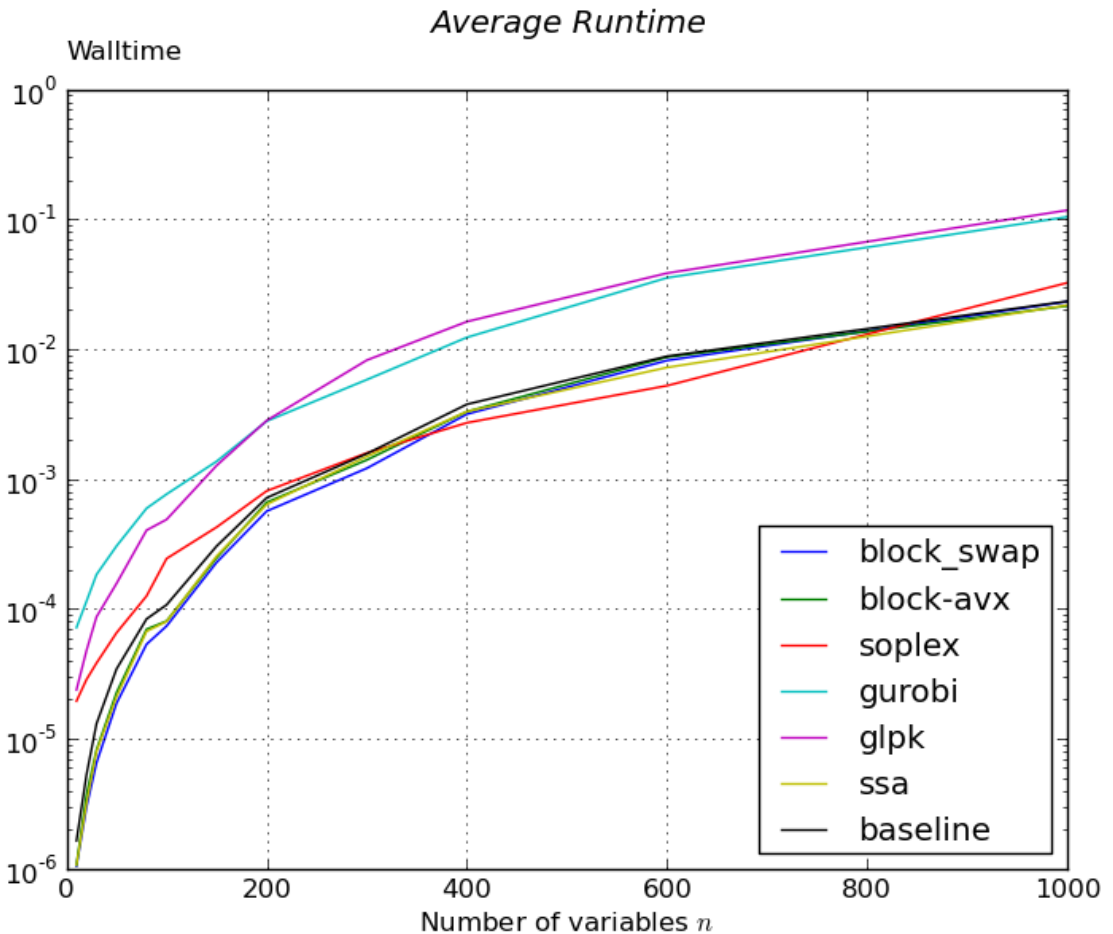


Figure : wall time comparison

Roofline

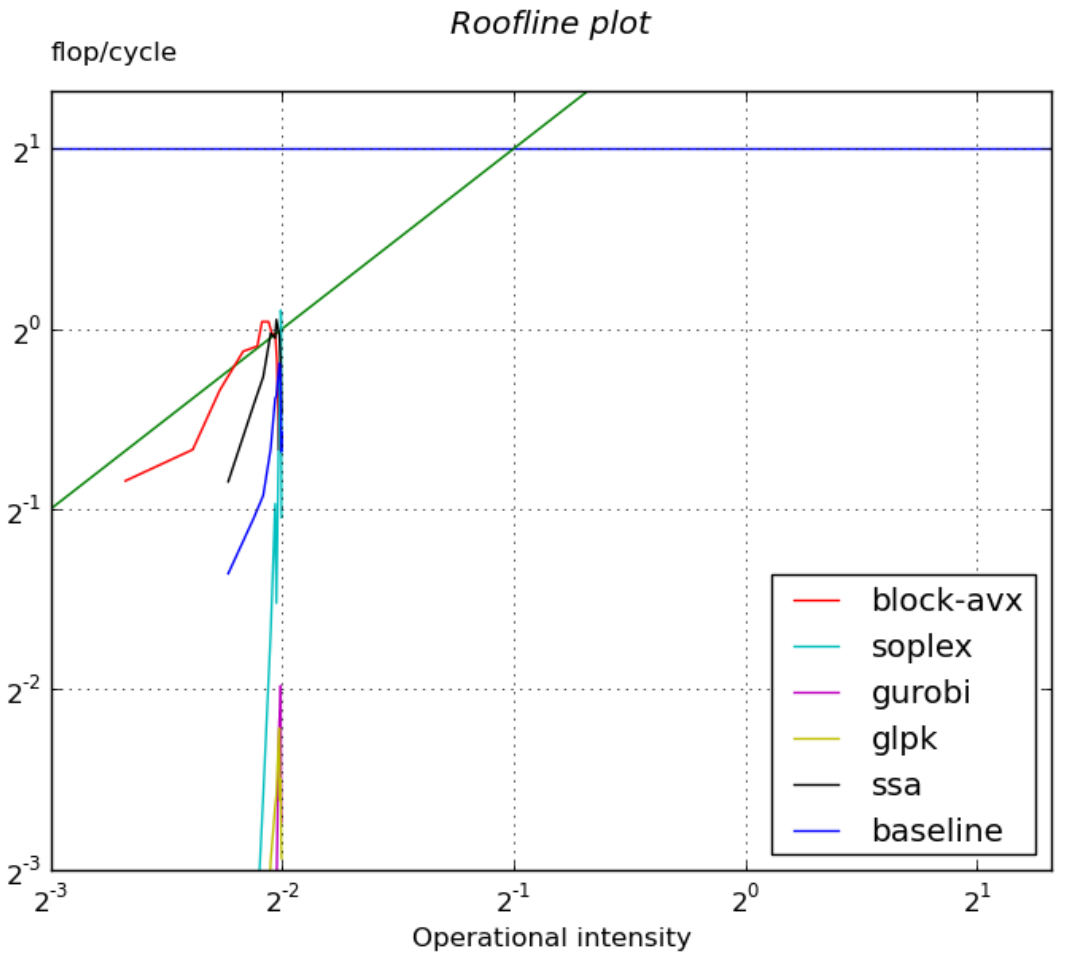
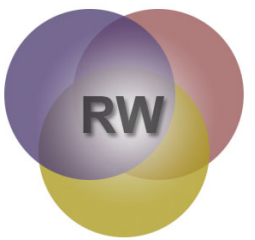


Figure : roofline



Profiling (gprof)

Memory load lines use most time

```
7.69      0.02     block_swap.hpp:122     T la1 = tabp[i*width+j];
3.85      0.01     block_swap.hpp:123     T la2 = tabp[i*width+j+1];
3.85      0.01     block_swap.hpp:124     T la3 = tabp[i*width+j+2];
3.85      0.01     block_swap.hpp:125     T la4 = tabp[i*width+j+3];
[... x16 ... ]
```