Matrix Balancing in L_p norms: A New Analysis of Osborne’s Iteration

Yuval Rabani
The Hebrew University of Jerusalem

joint work with:
Rafail Ostrovsky and Arman Yousefi (UCLA)

Matrix Balancing

- $n \times n$ real matrix $A = (a_{ij})_{i,j=1,\ldots,n}$
- it is balanced in $\|\cdot\|$ iff $\forall i$, $\|a_{i\cdot}\| = \|a_{\cdot i}\|$
- $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ balances A iff $D\,A\,D^{-1}$ is balanced ($a_{i\cdot}$ scaled by d_i, $a_{\cdot i}$ scaled by $1/d_i$)
- diagonal entries, signs of entries don't matter; w.l.o.g. A is nonnegative, has all-0s diagonal.

Osborne's Iteration

- balance index i: scale $a_{i\cdot}$ by $\sqrt{\|a_{\cdot i}\| / \|a_{i\cdot}\|}$ and $a_{\cdot i}$ by $\sqrt{\|a_{i\cdot}\| / \|a_{\cdot i}\|}$
- repeat (round-robin) until the matrix is balanced
- [Osborne 1960] the L_2 norm iteration converges to a balanced matrix
- [Parlett & Reinsch 1969] same iteration for any L_p norm
- $\mathrm{diag}(d_1, d_2, \ldots, d_n)$ balances A in the L_p norm iff $\mathrm{diag}(d_1^p, d_2^p, \ldots, d_n^p)$ balances (a_{ij}^p) in the L_1 norm
- [Grad 1971] convergence of the L_1 norm iteration

Why Balance?

- A and $D A D^{-1}$ have the same eigenvalues
- eigenvalue computations on unbalanced matrices are numerically unstable
- [Osborne 1960] if D balances A in L_2 then it minimizes the Frobenius norm of $D A D^{-1}$
- the numerical stability of eigenvalue computations depends on the Frobenius norm of the matrix
- Osborne's iteration is implemented in almost all numerical linear algebra software: MATLAB, LAPACK, EISPACK

Balanceable Matrices

- A is balaceable iff ∃D s.t. $D A D^{-1}$ is balanced
- G_A - weighted digraph on $\{1,2,\ldots,n\}$
  $(i,j) \in E(G_A)$ iff $a_{ij} > 0$, weight is a_{ij}
- if A is balanced in the L_1 norm, then the weights form a valid circulation in G_A
- [Eaves et al. 1985] A is balanceable iff G_A is strongly connected
- [Hartfiel 1971] D is unique up to uniform scaling

Balancing in L_1

[Kalantari-Khachiyan-Shokoufandeh 1997]

- def: A is ε-balanced if $\sqrt{\sum_i (||a_{\cdot i}|| - ||a_{i\cdot}||)^2} / \sum_{ij} a_{ij} \leq \varepsilon$
- Ellipsoid based algorithm, $O(n^4 \log(n \log \alpha / \varepsilon))$ arithmetic operations
  $\alpha = \sum_{ij} a_{ij} / a_{min}$ ($a_{min} = \min\{a_{ij}: a_{ij} > 0\}$)
- [Kalantari et al. 1997]: $d > 0$ minimizes $F(d) = \sum_{ij} a_{ij} (d_i / d_j)$ iff $D = \mathrm{diag}(d)$ balances $A = (a_{ij})$ in the L_1 norm
- minimize $f(x) = f_A(x) = \sum_{ij} a_{ij} \exp(x_i - x_j)$ is an unconstrained convex program

Balancing in L_∞

- [Schneider & Schneider 1991] $O(n^4)$ algorithm
- [Young, Tarjan & Orlin 1991] improved $O(mn + n^2 \log n)$
  m = #arcs of G_A
- [Chen 1998] Osborne's iteration converges to a balanced matrix, $\Theta(n^3)$ iterations when G_A is a directed cycle
- [Schulman & Sinclair 2015]: a <u>variant</u> (different order) of Osborne's iteration converges in $O(n^3 \log(\rho n/\varepsilon))$ iterations to an ε-balanced matrix, ρ = initial imbalance
- stronger notion of approximation: $\max_i |\log(||a_{\cdot i}|| \,/\, ||a_{i \cdot}||)| \le \varepsilon$

Our Results

- We analyze the convergence rate of three natural variants of Osborne's L_1 iteration:

- original - $O(\varepsilon^{-2} n^2 \log \alpha)$ iterations; $O(\varepsilon^{-2} mn \log \alpha)$ arithmetic operations on $O(n \log \alpha)$-bit numbers

- greedy - K iterations; $O(m + K n \log n)$ arithmetic operations on $O(n \log \alpha)$-bit numbers; $K = \min\{\varepsilon^{-2} \log \alpha, \varepsilon^{-1} n^{3/2} \log(\alpha/\varepsilon)\}$

- random - $O(\varepsilon^{-2} \log \alpha)$ iterations; $O(m + \varepsilon^{-2} n \log \alpha)$ arithmetic operations on $O(\log(\alpha n/\varepsilon))$-bit numbers

- lower bound: $\Omega(1/\sqrt{\varepsilon})$, any variant

Some Observations

- recall $f(x) = \sum_{ij} a_{ij} e^{x_i - x_j} = |||\ A(x)\ |||_1$
  $A(x) = (a(x)_{ij}) = D\ A\ D^{-1}$ for $D = \mathrm{diag}(e^{x_1}, \ldots, e^{x_n})$
- Osborne's iteration = coordinate descent to find
  $x^* = \mathrm{argmin}\ f(x)$
- $\partial f(x) / \partial x_i = ||a(x)_{i\cdot}||_1 - ||a(x)_{\cdot i}||_1$
- $\mathrm{diag}(e^{x_1}, \ldots, e^{x_n})$ ε-balances A iff $||\nabla f(x)||_2 / f(x) \leq \varepsilon$

Some Lemmas

- if $A(x) \mapsto A(x')$ as a result of balancing i, then $f(x) - f(x') \geq (\partial f(x) / \partial x_i)^2 / 2(\|a(x)_{i\cdot}\| + \|a(x)_{\cdot i}\|)$

- if i maximizes the drop in potential, then $f(x) - f(x') \geq (\|\nabla f(x)\|_2)^2 / 4f(x) = f(x)\cdot(\|\nabla f(x)\|_2 / 2f(x))^2$

- the challenging lemma: $f(x) - f(x^*) \leq (n/2)\cdot\|\nabla f(x)\|_1$

Distance to Optimality

Lemma: $f(x) - f(x^*) \leq (n/2)\cdot||\nabla f(x)||_1$

Proof: W.l.o.g. $x = 0$ (so $A(x) = A$).

- Recall $||\nabla f(0)||_1 = \sum_i |\, ||a_{i\cdot}||_1 - ||a_{\cdot i}||_1 \,|$
- Put $S = \{i: ||a_{\cdot i}||_1 > ||a_{i\cdot}||_1\}$ and $T = \{i: ||a_{\cdot i}||_1 < ||a_{i\cdot}||_1\}$
- Form a circulation by adding arcs between S and T
- Total added weight = $\sum_{i\in S} (||a_{\cdot i}||_1 - ||a_{i\cdot}||_1) = ½\cdot||\nabla f(0)||_1$
- Remove flow cycles via new arcs (cycle ≤ n arcs)
- Remaining weight ≥ $f(0) + ½\cdot||\nabla f(0)||_1 - (n/2)\cdot||\nabla f(0)||_1$

(cont.)

<u>Claim</u>: remaining weight ≤ $f(x^*)$

- flow cycles: C_k of length n_k, weight α_k, $k=1,2,\ldots$
- $G_{A'}$ = graph of remaining weights $= \sum_k \alpha_k C_k$
- $f(x^*) = \sum_{ij} a_{ij} \exp(x_i^* - x_j^*) \geq \sum_{ij} a'_{ij} \exp(x_i^* - x_j^*)$

$$= \sum_{ij} \sum_{k:ij\in C_k} \alpha_k \exp(x_i^* - x_j^*) = \sum_k \sum_{ij\in C_k} \alpha_k \exp(x_i^* - x_j^*)$$

$$\geq \sum_k n_k \left(\prod_{ij\in C_k} \alpha_k \exp(x_i^* - x_j^*)\right)^{1/n_k} = \sum_k n_k \alpha_k = \sum_{ij} a'_{ij}$$

arithmetic-geometric mean inequality

∏ weights along cycle invariant to balancing

Greedy Balancing

Recall: $f(x) - f(x') \geq f(x) \cdot (\|\nabla f(x)\|_2 / 2f(x))^2$

if $A(x)$ is not ε-balanced, $\|\nabla f(x)\|_2 / f(x) > \varepsilon$

Analysis #1: $f(x') < (1 - \varepsilon^2/4) \cdot f(x)$

$f(0) = \sum_{ij} a_{ij}$ and $f(x^*) \geq a_{min}$

Analysis #2: $\|\nabla f(x)\|_1 \leq n^{1/2} \cdot \|\nabla f(x)\|_2$

$f(x) - f(x') \geq \|\nabla f(x)\|_1 \cdot \|\nabla f(x)\|_2 / (4n^{1/2} \cdot f(x))$

$> (\varepsilon / 2n^{3/2}) \cdot (f(x) - f(x^*))$

Other Variants

- original algorithms - requires analyzing a phase
- random order - i is chosen with probability $(||a(x)_{i\cdot}|| + ||a(x)_{\cdot i}||) / 2f(x)$

The Lower Bound

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 101\varepsilon & 0 \\ 0 & \varepsilon & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad D\,A\,D^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & \sqrt{101}\varepsilon & 0 \\ 0 & \sqrt{101}\varepsilon & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$D = \mathrm{diag}(1, 1, \sqrt{101}, \sqrt{101})$$

- in one iteration a_{32} / a_{23} grows by a factor $\leq \dfrac{1+70\sqrt{\varepsilon}}{1+\varepsilon}$
- if A' is ε-balanced then $a'_{32} / a'_{23} > \dfrac{1}{100}$

Concluding Remarks

- how many iterations are needed to get $\forall i$, $\max\{||a_{\cdot i}||, ||a_{i\cdot}||\} / \min\{||a_{\cdot i}||, ||a_{i\cdot}||\} \leq 1 + \varepsilon$?

- tight bounds in terms of ε - can we get a bound of $\tilde{O}(n)$ which is also tight in terms of ε ?

- a practically appealing heuristic with better dependence on ε ?