Row Sampling in Other Norms

Presenter: Richard Peng
Nov 10, 2015

- **DISCLAIMER:** These notes are not necessarily an accurate representation of what I said during the class. They are mostly what I intend to say, and have not been carefully edited.

- **Summary**
  - Sampling to preserve $p$-norms.
  - $p - norm$ concentration bounds.
  - Iterative transformation to nice setting.

- **Problem:** given $A \in \mathbb{R}^{n \times d}$, find $\tilde{A}$ s.t. $\|\tilde{A}x\|_p \approx \|Ax\|_p$ for all $x \in \mathbb{R}^d$.
  - Application: speed up $\ell_1$-regression.
  - Difference with $\ell_2$ row sampling:
    * $n \times 1$ matrix, one row 1, rest all $1/n$.
    * $\ell_2$ case: all other rows have leverage score about $1/n^2$, no samples.
    * $\ell_1$: important to keep the other rows.
  - This works on graphs:
    * $L^{(G)} \approx L^{(H)}$ implies $\sum_{uv} w_{uv}^{(G)} |x_u - x_v| \approx \sum_{uv} w_{uv}^{(H)} |x_u - x_v|$.
    * Difference: scaling, these terms are actually $\|W^{1/2}A\|_2$ and $\|WA\|_1$.

- **Wishlist for sampling probabilities $w$:**
  - Add up to something small.
  - Invariant under right multiplicatoin, $A$ same as $AM$.
  - Invariant under splitting: splitting row $i$ into 2 copies gives 2 rows with probability $w_i/2$.

- **What does work:** $w_i = \sqrt{a_i (a_iW^{-1}a_i^T)^{-1}a_i^T}$, 1-norm Lewis weights.

- **$p$-norm matrix concentration bound:**
  - Modified isotropic position: if $A$ is a matrix such that:
    * $A^T A = I$, and
all rows of $A$ have the same norm, $\frac{d}{n}$.

Then a uniform sample of $n' = O(d \log d \epsilon^{-2})$ rows, rescaled by $n/n'$ gives $\tilde{A} \approx_\epsilon A$.

$1 < p < 2$: extra factor of $\log \log d$, $p > 2$: $d^{p/2} \log d$.

- **Transformation to this setting:**
  - Split a row $a_i$ into $t w_i$ copies: should now have weight $\frac{1}{t}$ each.
  - Weight of each copy: $\frac{1}{t w_i} a_i$.

- **Normalize to $A^T A = I$:**
  - Quadratic form, after splitting:
    $$\tilde{A}^T \tilde{A} = \frac{1}{t} \sum_i \frac{1}{w_i} a_i^T a_i = A^T W^{-1} A.$$

  - Norm of a single row:
    $$\frac{1}{t w_i^{-2}} a_i M^{-1} a_i^T.$$

  - Should be $1/t$:
    $$w_i^2 = a_i (A^T W^{-1} A)^{-1} a_i$$

- **Connection to leverage scores:**
  - $w_i$ is the leverage score of $W^{-1/2} A$.
  - If we set $\alpha = 1$, $\sum_i w_i = d$, exactly same as Foster’s theorem.

- **Computing $w$:**
  - Iterative scheme: $w' \leftarrow \left( a_i (A^T W^{-1} A)^{-1} a_i \right)^{1/2}$.
  - Convergence: show if $w \approx_k w'$, $w'' \approx_{\sqrt{k}} w'$:
    * Composition: $A^T W^{-1} A \approx_k A^T W'^{-1} A$.
    * Invert: $(A^T W^{-1} A)^{-1} \approx_k (A^T W'^{-1} A)^{-1}$.
    * Apply to vector $a_i$, and taking square roots gives the $k^{1/2}$ new bound.