CMPSCI 711: More Advanced Algorithms

Section 1-5: Sparse Approximations and Algebraic Approximations

Andrew McGregor

Last Compiled: April 29, 2012

Sparse Recovery

- *Goal:* Find g such that $\|f - g\|_1$ is minimized subject to the constraint that g has at most k non-zero entries.
- Define $\text{Err}^k(f) = \min_{g:\|g\|_0 \leq k} \|f - g\|_1$
- Exercise: $\text{Err}^k(f) = \sum_{i \notin S} |f_i|$ where S are indices of k largest f_i
- Using $O(\epsilon^{-1} k \log n)$ space, we can find $\tilde{g}$ such that $\|\tilde{g}\|_0 \leq k$ and

$$\|\tilde{g} - f\|_1 \leq (1 + \epsilon)\text{Err}^k(f)$$

Count-Min Revisited

- Consider Count-Min sketch with depth $d = O(\log n)$, width $w = \frac{4k}{\epsilon}$
- For $i \in [n]$, let $\tilde{f}_i = c_{j,h_j(i)}$ for some row $j \in [d]$.
- Let $S = \{i_1, \ldots, i_k\}$ be the indices with maximum frequencies. Let A_i be the event that there doesn't exist $k \in S \setminus i$, with $h_j(i) = h_j(k)$.
- Then for $i \in [n]$,

$$\begin{aligned}
\mathbb{P}\left[|f_i - \tilde{f}_i| \geq \epsilon\frac{\text{Err}^k(f)}{k}\right] &= \mathbb{P}\left[\neg A_i\right] \times \mathbb{P}\left[|f_i - \tilde{f}_i| \geq \epsilon\frac{\text{Err}^k(f)}{k}|\neg A_i\right] + \\
&\qquad \mathbb{P}\left[A_i\right] \times \mathbb{P}\left[|f_i - \tilde{f}_i| \geq \epsilon\frac{\text{Err}^k(f)}{k}|A_i\right] \\
&\leq \mathbb{P}\left[\neg A_i\right] + \mathbb{P}\left[|f_i - \tilde{f}_i| \geq \epsilon\frac{\text{Err}^k(f)}{k}|A_i\right] \\
&\leq k/w + 1/4 < 1/2
\end{aligned}$$

- With high probability, all f_i are approximated up to error $\epsilon\text{Err}^k(f)/k$

Sparse Recovery Algorithm

- Consider a Count-Min sketch with depth d and width $w = 4k/\epsilon$
- Let $f' = (\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n)$ be frequency estimates where for all i

$$|f_i - \tilde{f}_i| \leq \epsilon \frac{\mathrm{Err}^k(f)}{k}$$

- Let $\tilde{g}$ be f' with all but the kth largest entries replaced by 0.
- *Lemma:* $\|\tilde{g} - f\|_1 \leq (1 + 3\epsilon)\mathrm{Err}^k(f)$

Proof of Lemma

- Let $S, T \subset [n]$ be indices corresponding to largest values of f_i and $\tilde{f}_i$.
- For a vector $x \in \mathbb{R}^n$ and $I \subset [n]$, write x_I as the vector formed by zeroing out all entries of x except for those indices in I.
- Then.

$$\begin{aligned}
\|f - f'_T\|_1 &\leq \|f - f_T\|_1 + \|f_T - f'_T\|_1 \\
&= \|f\|_1 - \|f_T\|_1 + \|f_T - f'_T\|_1 \\
&= \|f\|_1 - \|f'_T\|_1 + (\|f'_T\|_1 - \|f_T\|_1) + \|f_T - f'_T\|_1 \\
&\leq \|f\|_1 - \|f'_T\|_1 + 2\|f_T - f'_T\|_1 \\
&\leq \|f\|_1 - \|f'_S\|_1 + 2\|f_T - f'_T\|_1 \\
&\leq \|f\|_1 - \|f_S\|_1 + (\|f_S\|_1 - \|f'_S\|_1) + 2\|f_T - f'_T\|_1 \\
&\leq \|f - f_S\|_1 + \|f_S - f'_S\|_1 + 2\|f_T - f'_T\|_1 \\
&\leq \mathrm{Err}^k(f) + k\epsilon \mathrm{Err}^k(f)/k + 2k\epsilon \mathrm{Err}^k(f)/k \\
&= (1 + 3\epsilon)\mathrm{Err}^k(f)
\end{aligned}$$

Similar Result for ℓ_2

- *Goal:* Find g such that $\|f - g\|_2$ is minimized subject to the constraint that g has at most k non-zero entries.
- Define $\text{Err}_2^k(f) = \min_{g : \|g\|_0 \leq k} \|f - g\|_2^2$
- Using $O(\epsilon^{-2} k \log n)$ space, we can find $\tilde{g}$ such that $\|\tilde{g}\|_0 \leq k$ and

$$\|\tilde{g} - f\|_2^2 \leq (1 + \epsilon)\text{Err}_2^k(f)$$

Outline

Haar Wavelets

- *Defn:* For $n = 8$, Haar Wavelet basis consists of rows of the matrix.

$$M = \begin{pmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} \\ \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0, & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

  and for $n = 2^r$, the construction generalizes in the natural way.
- Note that the basis is orthonormal and $MM^T = M^T M = I$. Hence, any signal $f \in \mathbb{R}^n$ can be expressed in the Haar basis.

Sparse Representation in Haar Basis

- Let $\mathcal{H} = \{\phi_1, \ldots, \phi_n\}$ be the Haar basis.
- *Goal:* Find g that minimizes $\|f - g\|_2$ subject to the constraint that g can be expressed as the sum of at most k Haar basis vectors, i.e., g is k-sparse in the Haar basis.
- Write $f = \sum_i \lambda_i \phi_i$ where $\lambda_i = \phi_i \cdot f$.
- Suppose $g = \sum_{i \in I} \mu_i \phi_i$ for some $I \subset [n]$ of size at most k.
- Then
$$\|f - g\|_2^2 = \sum_{i \in I} (\lambda_i - \mu_i)^2 + \sum_{i \notin I} \lambda_i^2$$
- Hence, we want to find k values of i that maximize $\mu_i = \phi_i \cdot f$.

Time Series Model

- Suppose coordinations of f are presented in order $\langle f_1, f_2, \ldots, f_n \rangle$. This is called the *time-series model*.
- Can easily compute $\mu_i = \phi_i \cdot f$
- At any given time,
  - We've calculated μ_i exactly for some $i \in A$
  - We've calculated μ_i partially for some $i \in B$
  - We haven't started computing μ_i for $i \notin A \cup B$
- *Lemma:* The size of B is at most $\log_2 n$.
- *Algorithm:* Maintain only the k largest values of μ_i for $i \in A$.
- We find the optimal k term representation in $O(k + \log n)$ space.

General Update Model

- Can express goal in terms of standard basis...
- Because M is unitary,

$$\|f-g\|_2^2 = (f-g)^T(f-g) = (f-g)^T M^T M(f-g) = \|Mf-Mg\|_2^2$$

  and g is k-sparse in Haar basis iff Mg is k-sparse in standard basis.
- Hence, finding best g is same as finding $h = Mg$ with $\|h\|_0 \leq k$ that minimizes $\|Mf - h\|_2$
- Using Count-Sketch algorithm, can find $\tilde{h}$ with $\|\tilde{h}\|_0 \leq k$ such that

$$\begin{aligned}
\|Mf - \tilde{h}\|_2^2 &\leq (1+\epsilon) \min_{h:h \text{ is } k\text{-sparse in standard basis}} \|Mf - h\|_2^2 \\
&= (1+\epsilon) \min_{g:g \text{ is } k\text{-sparse in Haar basis}} \|Mf - Mg\|_2^2 \\
&= (1+\epsilon) \min_{g:g \text{ is } k\text{-sparse in Haar basis}} \|f - g\|_2^2
\end{aligned}$$