6.893: Algorithms and Signal Processing

Lecture 11:
Structured Sparsity Models
Structured sparsity models

• There is more to modeling sparsity than just specifying the value of $k$
• Examples:
  – k-histogram (subsumes k-sparsity)
  – tree sparsity (is subsumed by k-sparsity)
Piece-wise constant functions (histograms)

• A function $h$ from $\{0..n-1\}$ to $\mathbb{R}$ defined by
  - $k$ disjoint intervals $I_1...I_k$ covering $\{0..n-1\}$
  - $k$ values $v_1...v_k$

  as $h(i)=v_j$ for $I_j$ containing $i$

• Given a vector $x$, we want to find $h^*$ such that

\[
\|h^*-x\|_2^2 \leq A \min_{k\text{-histogram } h} \|h-x\|_2^2
\]

where $h^*$ is a $B_k$-histogram
## Algorithms for k-histograms

<table>
<thead>
<tr>
<th>Running time</th>
<th>[A,B]-approximation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O(kn^2)$</td>
<td>exact</td>
<td>JKMPST’98</td>
</tr>
<tr>
<td>$O(n)$</td>
<td>$[1, \log n + 1]$</td>
<td>Previous lecture</td>
</tr>
<tr>
<td>$O(n^2)$</td>
<td>$[1, 2]$</td>
<td>JKMPST’98</td>
</tr>
<tr>
<td>$O(n^*)$</td>
<td>$[O(1), O(1)]$</td>
<td>JKMPST’98</td>
</tr>
<tr>
<td>$O(nk^2)$</td>
<td>$[1+\varepsilon, 1]$</td>
<td>GKS’01</td>
</tr>
</tbody>
</table>

* Assuming the optimal error value is known up to a constant factor
Quadratic-time algorithm

• \( \text{OPT}[t,i]=\text{minimum cost of } t \)-histogram for \( x_1 \ldots x_i \)

• Recursive formula:
  \[
  \text{OPT}[t,i]=\min_{j<i}\text{OPT}[t-1,j]+\text{Var}[j+1 \ldots i]
  \]
  where
  \[
  \text{Var}[j+1 \ldots i] = \sum_{t=j+1}^{i} x_t^2 - (\sum_{t=j+1}^{i} x_t/(i-j))^2
  \]

• Space: \( O(kn) \)

• Time: \( O(kn^2) \)
Linear-time algorithm

• Lemma: Suppose that x has a k-histogram h with error E. Then there is a k’-histogram h’, k’=O(k), such that the error in each interval of h’ is at most E’=E/k.

• Proof: consider each interval I of h
  – If Var[x_I]≤E/k, then we are fine
  – If Var[x_I]>E/k, then split I into t intervals for t=O(Var[x_I] / (E/k))

• It suffices to find k’-histogram where each interval has variance E’
Linear-time algorithm ctd.

• Algorithm:
  – range = [1…n]
  – While range = [a..n] is nonempty
    • Let b be the largest index s.t. \( \text{Var}(x_{[a..b]}) \leq E' \)
    • Output interval [a…b]
    • range = [b+1…n]

• Correctness:
  – Suffices to show that the algorithm takes at most \( k' \) steps
  – Consider any step of the algorithm, with range = [a…n]
  – Let I = [c…d] be the interval in the optimal \( h \) that contains a
  – Observation: the algorithm will compute \( b \geq d \)
  – Therefore, each step of the algorithm deals with at most one interval of \( h \)
  – At most \( k' \) steps
Tree-sparsity

- Sparse signals whose large coefficients can be arranged in the form of a rooted, connected tree
Tree sparse approximation

• Given a signal $x$, compute the optimal (exact) \textbf{tree-sparse projection} of $x$, i.e., solve

$$\min_{|\Omega| \leq k, \Omega \text{ tree}} \|x - x_{\Omega}\|_2^2$$
Tree sparse approximation

- Given a signal $x$, compute the optimal (exact) tree-sparse projection of $x$, i.e., solve

$$\max_{|\Omega| \leq k, \Omega \text{ tree}} \|x_{\Omega}\|_2^2$$
## Algorithms for tree sparse approximation

<table>
<thead>
<tr>
<th>Running time</th>
<th>Approximation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O(n^2)$</td>
<td>exact</td>
<td>Bohanec-Bratko ‘94</td>
</tr>
<tr>
<td>$O(nk)$</td>
<td>exact</td>
<td>Cartis-Thompson ‘13</td>
</tr>
<tr>
<td>$O(n \log n)$</td>
<td>$[O(1),O(1)]$</td>
<td>Hegde-Indyk-Schmidt’14</td>
</tr>
</tbody>
</table>
Exact algorithm

• $\text{OPT}[t,i] = \max_{|\Omega| \leq t} \ |x_{\Omega}|_2^2$

• Recurrence:
  – If $t>0$ then
    \[ \text{OPT}[t,i] = \max_{0 \leq s \leq t-1} \text{OPT}[s,\text{left}(i)] + \text{OPT}[t-1-s,\text{right}(i)] + x_i^2 \]
  – If $t=0$ then $\text{OPT}[0,i]=0$

• Space:
  – On level $i$: $2^i \cdot n/2^i = n$
  – Total: $n \log n$

• Running time:
  – On level $i$ s.t. $2^i < k$: $n \cdot 2^i$
  – On level $i$ s.t. $2^i \approx k$: $nk$
  – On level $i+1$: $nk/2$
  – ...
  – Total: $O(nk)$
Approximate algorithm

- We want to approximate:
  \[ \min_{|\Omega|\leq k, \ \Omega \text{ tree}} ||x-x_\Omega||_2^2 \]

- Perform a Lagrange relaxation of the sparsity constraint:
  \[ \min_{\Omega \text{ tree}} ||x-x_\Omega||_2^2 + \lambda|\Omega| \]

- Can be solved using a simple dynamic program [Donoho ’97]:
  \[ \text{OPT}[i] = \min[x^2_{\text{subtree}(i)}, \ \lambda + \text{OPT}[\text{left}(i)] + \text{OPT}[\text{right}(i)] \]
How to choose $\lambda$?

- Pareto analysis: examine the solution curve as a function of $\lambda$.
  
- Via convexity of the curve, we can always obtain:
  
  - Either: $(\leq 2k, \leq \text{OPT})$
  - Or: $(\leq k, \leq 2\text{OPT})$

- Altogether: we get a tail guarantee in time roughly $O(n \log \text{RANGE})$.

- Can be improved to $n \log n$. 

![Graph with Pareto curve](graph.png)