Complementary material

Sparse representation

orthogonal matching pursuit (OMP) algorithm

Orthogonal Matching Pursuit (OMP)

- Like MP, OMP also applies the SFS principle to select the bases, but the difference is that the corresponding coefficients (c) are allowed to be changed in the iterations.
- Hence, OMP can approximate the input signal x better, because the coefficients are further optimized. However, OMP is slower.

OMP algorithm principle

- First, like MP, finding the first basis by

$$b_{s_1} = \underset{b_i,\, i=1\ldots m}{\operatorname{argmax}} \left| b_i^T x \right|$$

Then, assume l bases have been found, find the $(l+1)$-th basis according to the following principle.

OMP algorithm principle

- Assume l bases $b_{s_1}, b_{s_2}, \dots, b_{s_l}$ have been found, then the $(l+1)$-th basis and all the $l+1$ coefficients are found as

$$\left[b_{s_{l+1}}, c_{\Lambda_{l+1}}\right] = \underset{b \in B_{\overline{\Lambda}_l};\, u}{\operatorname{argmin}} \left\| x - \left[B_{\Lambda_l} \quad b\right] u \right\|^2, \qquad (1)$$

where $\Lambda_l = \{s_1, s_2, \dots, s_l\}$ is the set of indices already selected, and B_{Λ_l} is the submatrix consisting of the associated columns of B.

$\overline{\Lambda}_l = \{1, \dots, m\} \backslash \{s_1, s_2, \dots, s_l\}$ is the complementary set of Λ_l, b is an $n \times 1$ vector, and u is an $(l+1) \times 1$ vector.

OMP algorithm principle (cont.)

- When b is fixed, the solution of (1) is obtainable by the pseudo inverse,

$$\hat{u}(b) = \left(M^T M\right)^{-1} M^T x,$$

where $M = \begin{bmatrix} B_{\Lambda_l} & b \end{bmatrix}$.

- So, (1) can be solved by calculating $\hat{u}(b)$ for all the remaining bases b, $b \in B_{\overline{\Lambda}_l}$,

$$b_{s_{l+1}} = \underset{b \in B_{\overline{\Lambda}_l}}{\mathrm{arg}min} \left\| x - \begin{bmatrix} B_{\Lambda_l} & b \end{bmatrix} \hat{u}(b) \right\|^2,$$

and the optimal coefficients are obtained as

$$c_{\Lambda_{l+1}} = \hat{u}(b_{s_{l+1}}).$$

- Finally, the stopping criterion is the same as that in MP.

Redundancy of computation

- However, in the above method, the pseudo inverse $\left(M^T M\right)^{-1} M^T$ is computed $m - l$ times in the l-th iteration.
- The common OMP algorithm adopts a residual-signal representation like MP, and the pseudo inverse can be computed **only once** in each iteration; a more efficient algorithm is yieldable as follows.

More efficient computation

- In the l-th iteration, we have already obtained the optimal bases B_{Λ_l} and coefficients c_{Λ_l} so far.
- Assume that $r = x - B_{\Lambda_l} c_{\Lambda_l}$. (residual signal)
- **Property**

– In the $(l+1)$-th iteration, the optimal basis in eq. (1),

$$- \quad \left[b_{s_{l+1}}, \sim\right] = \underset{b \in B_{\bar{\Lambda}_l}; u}{\operatorname{argmin}} \left\| x - \begin{bmatrix} B_{\Lambda_l} & b \end{bmatrix} u \right\|^2,$$

– can be equivalently and more easily found as

$$- \quad b_{s_{l+1}} = \underset{b \in B_{\bar{\Lambda}_l}}{\operatorname{argmax}} \left| b^T r \right|.$$

- Hence, we can use the inner product $b^T r$ instead of pseudo inverse in the computation. $\rightarrow$ More efficient.

OMP Algorithm

- **Initialization**: let the residue signal be $r_0 = x$; set $\Lambda_0 = \phi$ (empty set), the iteration count $l = 1$.
- **Iteration**:

  Find the index s_l that solves the problem,

  $$b_{s_l} = \underset{b \in B_{\bar{\Lambda}_{l-1}}}{\mathrm{argm}ax} \left| b^T r_{l-1} \right|.$$

  Let $\Lambda_l = \Lambda_{l-1} \cup \{s_l\}$.

  Solve the following least squares problem,

  $$c_{\Lambda_l} = \underset{c_t}{\mathrm{argmin}} \left\| x - B_{\Lambda_l} c_t \right\|^2,$$

  (the solution is $c_{\Lambda_l} = \left(B_{\Lambda_l}^t B_{\Lambda_l} \right)^{-1} B_{\Lambda_l}^t x$)

  Let $r_l = x - B_{\Lambda_l} c_{\Lambda_l}$,
- **Stopping criterion:** same as MP.

OMP

- OMP has been widely used to solve the sparse-representation problem.
  - Another widely used method is basis-pursuit or Lasso that solves the l_1-norm problem that is a convex relaxation of the l_0-norm problem.
- It also serves as a component for some dictionary learning methods, such as KSVD.
- OMP has a theoretical guarantee to get the exact solution under some conditions, which will be introduced in the following.

Mutual Incoherence

- Given a dictionary matrix $B \in R^{n \times m}$, where the columns of B are unit-length bases, i.e., $\|b_i\|^2 = 1$ with b_i the i-th column of B.
- The coherence parameter μ of a matrix B is defined as

$$\mu = \max_{i \neq j} \left| b_i^T b_j \right|,$$

i.e., the coherence is the maximal absolute value of the inner products between two different bases.

- Eg., when $m = n$ and B is an orthonormal matrix, then $\mu = 0$.
- The range of $\mu \in [0,1]$. The smallest is μ, the closer is B to an orthonormal matrix.

k-sparse signal

- Assume that a signal x is exactly represented as a linear combination of k bases in B. That is,

$$x = Bc, \ \|c\|_0 = k,$$

then x is called a ***k*-sparse signal** (with respect to the bases B).

- If x is exactly k-sparse, can we find the reconstruction coefficients c by OMP?

Mutual incoherence (MI) condition

- To recover the coefficients for a k-sparse signal represented by the bases matrix B, if the coherence parameter of B satisfies

$$\mu < \frac{1}{2k-1},$$

it is called that B satisfies the mutual incoherence (MI) condition.

- **Property**: if B satisfies the MI condition, then the coefficients c of any k-sparse signal $x = Bc$ can be exactly recovered by using the OMP algorithm.

OMP property

- If the MI condition is not satisfied, i.e., $\mu \geq \frac{1}{2k-1}$, then OMP cannot ensure finding the exact k-sparse solution.
- However, in this case, OMP still finds an approximate sparse solution instead.
- OMP, as a greedy search approach in basis finding, is thus an effective method with some theoretical guarantee.

Is the MI condition easy to achieve? squared vs. over-complete matrices

- For a dictionary where the the basis matrix B is a square matrix, there is a lot of room to make the bases “mutually incoherent.”
  - Eg., if B is an orthonormal matrix, then the inner products of two different bases are zero. The bases are “extremely incoherent” in this case.
- However, when $B \in R^{n \times m}$ is over-complete, i.e., a ‘fat’ matrix with $m > n$, the bases are linear dependent in general, and they cannot be orthonormal anymore.

Mutual incoherence in over-complete dictionary

- Thus, the coherence parameter μ is considered a measurement on "how close is it between an over-complete matrix and an orthonormal matrix?" (with the columns of the matrix normalized to unit-length).
  - It acts like a counterpart of "ortohnormality" for a 'fat' matrix.
  - The lower is μ, the more incoherent is the matrix.
  - Intuitively, we can image that the 'fatter' is the matrix, the largest is the lower bound of μ.

Incoherence for constructing over-complete dictionaries

- When choosing, constructing, or learning an over-complete dictionary, sometimes we hope that it is as mutually incoherent as possible.
  - The coherence parameter reflects the correlation between two bases.
  - When any of the two bases are close, the dictionary then contains 'redundancy' because removing one of them or combining them could result in a similar dictionary.
  - We often hope that the dictionary is as **'non-redundant'** as possible.

Lower bound of the coherence parameter

- An important theoretical problem then occurs. What is the lower bound of the coherence parameter μ for a $n \times m$ $(m \geq n)$ matrix?
- The answer is a special case of the **Welch bound**, stated as follows:
- Property [Welch bound]: Given a bases matrix $B \in C^{n \times m}$ $(m \geq n)$, its coherence parameter, namely μ_B, satisfies that

$$\mu_B \geq \sqrt{\frac{m-n}{n(m-1)}}.$$

Example of the Welch bound

- Eg, for there are $m = 16$ bases in an $n = 8$ dimensional space, the Welch bound is 0.258199, i.e. the minimal angle between two bases is $\cos^{-1}(0.258199) = 75.036783^{\circ}$ when the bound is achieved.
- For $m = 400$ and $n = 100$, the Welch bound is 0.086711, and the minimal angel is 85.025579° when the bound is achieved.