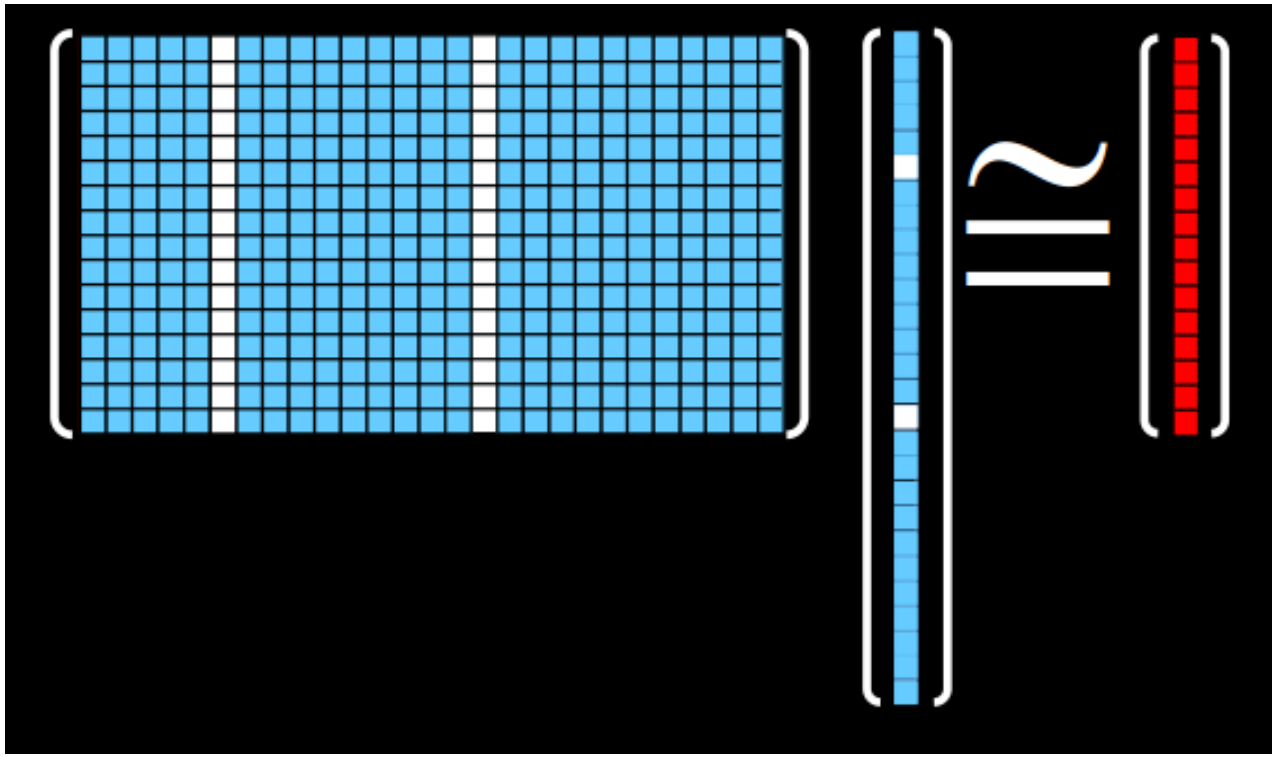
SGN 21006 Advanced Signal Processing
Project bonus: Sparse model estimation

Ioan Tabus

Department of Signal Processing
Tampere University of Technology
Finland

Sparse models



- Initial problem: solve a system of equations such that the solution has only few non-zero elements.
- In each equation equality has to be enforced to be exact or, is allowed to be an approximation, within a certain bound.
- If an exact solution of sparsity k exists, there are some sparse modelling methods that are guranteed to find the solution, under some conditions on k and on the system matrix.

Sparse modelling problem (1)

- Given the input to the adaptive filter $u(1), u(2), \ldots, u(N)$ and the desired signal $d(1), \ldots, d(N)$
- Consider the filter impulse response $\underline{w} = [w_0, w_1, \ldots, w_{M-1}]^T$, where only a small number of coefficients are non-zero. Such a vector is called sparse vector, the number of non-zero elements, k, is called sparsity, and is also denoted as $k = \|\underline{w}\|_0$.
- The model equation at time n is

$$d(n) = w_0 u(n) + w_1 u(n-1) + \ldots w_{M-1} u(n-(M-1)) + e(n)$$

where $e(n)$ is the model residual.

- By writing the model equation for $n = 1, 2, \ldots, N$ (using any data-windowing method) and denoting the data matrix as A, one gets the system of equations

$$\underline{d} \quad = \quad A\underline{w} + \underline{e} \tag{1}$$

as in the LS lecture, where the LS solution $\underline{\hat{w}}$, minimizing the sum of squares $\underline{e}^T \underline{e} = \sum_{n=1}^{N} e(n)^2$, is given by the solution of the system

$$(A^T A)\underline{\hat{w}} \quad = \quad A^T \underline{d}$$

- This solution is called also a ℓ_2-norm solution, since it minimizes the ℓ_2 norm of the error vector $\underline{e}$, defined as $\|\underline{e}\|_2 = \underline{e}^T \underline{e} = \sum_{n=1}^{N} e(n)^2$. This solution, in general, is not sparse.

Sparse modelling problem (2)

- One formulation of the sparse estimation problem is

$$\min \|\underline{w}\|_0 \quad \text{such that} \quad \|\underline{e}\|_2 < \delta \tag{2}$$

  i.e., find a sparse model $\underline{w}$, having the smallest number $k = \|\underline{w}\|_0$ of non-zero parameters, such that the norm of residuals is not larger than a given bound δ.

- One way to solve the above problem is by using greedy algorithms, like OMP or OOMP, see next pages.

Greedy algorithms for sparse design

- We redenote the variables in $\underline{d} = A\underline{w} + \underline{e}$ (Eq.1) in an obvious way to get

$$\mathbf{y} = \mathbf{D}\mathbf{\Theta} + \varepsilon$$

  ($\mathbf{y}$ is the old $\underline{d}$, $\mathbf{D}$ is the old A, and ε is the old $\underline{e}$).
- We denote the number of input and desired input datapoints as n (i.e., there are n rows in $\mathbf{y}$, in $\mathbf{D}$, and in ε).
- The greedy sparse design starts at stage 1 by designing the best predictor having a single non-zero coefficient in $\mathbf{\Theta}$, and then continues adding one new non-zero parameter at each stage.

Greedy algorithms: first stage

- The matrix $\mathbf{D}$ is denoted $[\mathbf{d}_1\ \mathbf{d}_2 \ldots \mathbf{d}_m]$.
- If the saprsity of the solution is 1, let assume that the ith element of $\mathbf{\Theta}$ is non-zero, and denote $\theta = \mathbf{\Theta}_i$. Then the model is

$$\mathbf{y} = \mathbf{D}\mathbf{\Theta} + \boldsymbol{\varepsilon} = \mathbf{d}_i\theta + \boldsymbol{\varepsilon}$$

This LS problem can be solved immediately

$$\hat{\theta} \quad = \quad (\mathbf{d}_i^T\mathbf{d}_i)^{-1}(\mathbf{d}_i^T\mathbf{y}) \tag{3}$$

and leads to a LS sum of squares

$$\begin{aligned} \|\varepsilon\|_2 \quad &= \quad \boldsymbol{\varepsilon}^T\varepsilon = \left(\mathbf{y} - \mathbf{d}_i\hat{\theta}\right)^T\left(\mathbf{y} - \mathbf{d}_i\hat{\theta}\right) \\ &= \quad \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{d}_i\hat{\theta} + \mathbf{d}_i^T\mathbf{d}_i\hat{\theta}^2 = \mathbf{y}^T\mathbf{y} - \frac{\left(\mathbf{y}^T\mathbf{d}_i\right)^2}{\mathbf{d}_i^T\mathbf{d}_i} \end{aligned} \tag{4}$$

The solution (3) and the corresponding criterion $\|\varepsilon\|_2$ from (4) are computed for each of the columns $\mathbf{d}_i$, and that i^* for which $\|\varepsilon\|_2$ is minimum is picked as the regressor column for Stage 1. Then we procced to next iterative stages, to pick the second element, and then third, see next algorithms.

Greedy algorithms for sparse design (2)

- We denote by $\mathbf{\Theta}$ a sparse predictor, where the non-zero elements with indices i in the support set $\mathcal{I} \subset \{1, \ldots, m\}$ are nonzero. The non-zero elements $\mathbf{\Theta}_i$, for $i \in \mathcal{I}$, form the vector denoted $\boldsymbol{\theta} = \mathbf{\Theta}_{\mathcal{I}}$ and the model prediction can be written as $\hat{\mathbf{y}} = \mathbf{D}\mathbf{\Theta} = \mathbf{D}_{\mathcal{I}}\boldsymbol{\theta}$, where $\mathbf{D}_{\mathcal{I}}$ is selecting from the matrix $\mathbf{D}$ only the columns with indices from $\mathcal{I}$.
- One instance of the sparse design problem is defined by the $n \times m$-matrix $\mathbf{D}$ and the n-vector $\mathbf{y}$ and the task is to find a (sub)optimal sparse predictor with support $\mathcal{I}_{k^*} \subset \{1, \ldots, m\}$ and the nonzero predictor parameters $\boldsymbol{\theta} \subset \mathbb{R}^{k^*}$.
- The prediction residuals are

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{D}_{\mathcal{I}_{k^*}}\boldsymbol{\theta}.$$

Two greedy algorithms for sparse design

- We consider two algorithms: orthogonal matching pursuit (OMP), optimized orthogonal matching pursuit (OOMP).
- They belong to the greedy class of algorithms and have similar algorithmic structure, where the solutions are constructed recursively in the sparsity k of the predictor, i.e., in the number of nonzero elements of the predictor.
- The solutions are constructed for all $k = 1, 2, \ldots, K$ where K is the maximum number of regressors variables allowed by the user in the model, out of all m existing regressors.
- In a greedy algorithm the set $\{1, \ldots, m\}$ of all regressor indices is split at each stage k into the chosen regressors set, $\mathcal{I}_k$ and available regressors set $\mathcal{A}_k$.
- The algorithm progresses from stage $k-1$ by enlarging the predictor support $\mathcal{I}_{k-1}$ (which was found the best at sparsity $k-1$) with a new regressor, to obtain the support $\mathcal{I}_k$, having cardinality k.

Two greedy algorithms for sparse design

- After the inner loop is terminated, the best index i^* for extending the support is found and used to update the support $\mathcal{I}_k$ and the set $\mathcal{A}_k$ of available indices for next outer loop stage.
- For the selected support $\mathcal{I}_k$ the LS solution $\boldsymbol{\theta}_k$ is computed at line 15, and the residual $\mathbf{r}_k$ corresponding to it is computed at line 16 (that is strictly needed only for the OMP algorithm).

Two greedy algorithms for sparse design, OMP and OOMP

Input: $\mathbf{y}$, $\mathbf{D}$
Initialize: $\mathcal{I}_0 = \emptyset$; $\mathcal{A}_0 = \{1, \ldots, m\}$
// Outer loop
for $k = 1$ to K **do**
  // Inner loop
  for $i \in \mathcal{A}_{k-1}$ **do**
    if Algorithm $=$ OOMP **then**
      $J_i = \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{D}_{\mathcal{I}_{k-1} \cup i} \boldsymbol{\theta}\|^2$ *// (solve a LS problem)*
    else if Algorithm $=$ OMP **then**
      $J_i = \|\mathbf{r}_{k-1}\|^2 - \frac{(\mathbf{r}_{k-1}^T \mathbf{D}_i)^2}{\mathbf{D}_i^T \mathbf{D}_i}$
    end if
  end for
  $i^* = \arg\min_i J_i$
  $\mathcal{I}_k = \mathcal{I}_{k-1} \cup i^*$
  $\mathcal{A}_k = \mathcal{A}_{k-1} \setminus i^*$
  $\boldsymbol{\theta}_k = \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{D}_{\mathcal{I}_k} \boldsymbol{\theta}\|^2$
  $\mathbf{r}_k = \mathbf{y} - \mathbf{D}_{\mathcal{I}_k} \boldsymbol{\theta}_k$
end for

The optimized orthogonal matching pursuit (OOMP) algorithm

- The algorithm OOMP evaluates at each inner iteration of the Algorithm (line 7) the solution of a full LS problem, resulting in the minimized LS criterion

$$J_i = \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{D}_{\mathcal{I}_{k-1} \cup i} \boldsymbol{\theta}\|^2 \tag{5}$$

- The already computed $k-1$ long support $\mathcal{I}_{k-1}$ is tentatively expanded to include $i \in \mathcal{A}_{k-1}$, and the resulting LS criterion J_i reflects the ability of the predictor with support $\mathcal{I}_{k-1} \cup i$ to model the data, even in the case when the columns of $\mathbf{D}$ are strongly correlated.
- The index $i^* = \arg\min_{i \in \mathcal{A}_{k-1}} J_i$ is selected to do the expansion $\mathcal{I}_k = \mathcal{I}_{k-1} \cup i^*$.
- There are several fast versions of this algorithm, exploiting the recursive-in-order relations existing between the predictor minimizing $\|\mathbf{y} - \mathbf{D}_{\mathcal{I} \cup i} \boldsymbol{\theta}\|^2$ and the one minimizing $\|\mathbf{y} - \mathbf{D}_{\mathcal{I}} \boldsymbol{\theta}\|^2$.

The orthogonal matching pursuit (OMP) algorithm

- OMP essentially evaluates a scalar product at each inner iteration (line 9 in Algorithm), and can be seen as a speeded-up version of OOMP.
- The new regressor index being selected is the one which gives the maximum in

$$i^* = \arg \max_{i \in \mathcal{A}_{k-1}} \frac{(\mathbf{r}_{k-1}^T \mathbf{D}_i)^2}{\mathbf{D}_i^T \mathbf{D}_i} \tag{6}$$

where the residual vector at stage $k-1$ is $\mathbf{r}_{k-1} = \mathbf{y} - \mathbf{D}_{\mathcal{I}_{k-1}} \boldsymbol{\theta}^*_{k-1}$.

- The connection to (5) can be seen by noting that

$$\min_{\theta} \|(\mathbf{y} - \mathbf{D}_{\mathcal{I}_{k-1}} \boldsymbol{\theta}^*_{k-1}) - \mathbf{D}_i \theta\|^2 = \min_{\theta} \|(\mathbf{r}_{k-1} - \mathbf{D}_i \theta\|^2 \tag{7}$$

$$= \|\mathbf{r}_{k-1}\|^2 - \frac{(\mathbf{r}_{k-1}^T \mathbf{D}_i)^2}{\mathbf{D}_i^T \mathbf{D}_i} \tag{8}$$

which reveals that the simpler criterion in (6), equivalent to the one dimensional LS problem (7), assumes during the iterations in the inner loop that the predictor coefficients $\boldsymbol{\theta}^*_{k-1}$ corresponding to the regressors with indices in $\mathcal{I}_{k-1}$ are not changing significantly in the solution for $\mathcal{I}_{k-1} \cup i$.