1 Problem Formulation

We are solving the inverse problem defined by the following forward model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the measurement, $\mathbf{x} \in \mathbb{R}^N$ is the signal of interest that we want to reconstruct, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the measurement matrix, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 I)$ is additive noise. We assume that the signal to reconstruct is at most K-sparse, i.e. $||\mathbf{x}||_0 \leq K$. Additionally, $K < M < N$ and the entries of $\mathbf{A}$ are i.i.d. Gaussians.

We are now learning or selecting a CPWL function that reconstructs the signal:

$$\hat{\mathbf{x}} = f(\mathbf{y}) = \mathbf{B}(\mathbf{y})\mathbf{y}, \tag{2}$$

where $\mathbf{B}(\mathbf{y})$ is the linear operation performed by f for input $\mathbf{y}$. Note that $\mathbf{B}(\mathbf{y}) \in \mathbb{R}^{N\times M}$.

We aim to show that the number of unique matrices that are assigned by an optimal maximum a-posteriori estimator f is limited, i.e. the optimal number of linear projection regions is limited. Note that from a frequentist perspective, the method will end up being the exact same as maximum likelihood estimation.

2 Calculating the posterior

We are interested in maximizing the posterior probability $p(\mathbf{x}|\mathbf{y})$. This can be found through Bayes' rule as:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

The prior $p(\mathbf{x})$ is a uniform prior over all possible reconstructions $\mathbf{x}$ that are of maximum cardinality K and is zero for all other possibilities. So we can also write it as:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \propto \begin{cases} p(\mathbf{y}|\mathbf{x}) & \text{if } |S| \leq K \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where $S = \mathrm{supp}(\mathbf{x})$ denotes the support of $\mathbf{x}$. We are interested in finding the MAP estimate, so given the proportional to statements, the MAP becomes:

$$\hat{\mathbf{x}} = \arg\max_{\substack{\mathbf{x} \\ |S|\leq K}} p(\mathbf{y} \mid \mathbf{x}). \tag{5}$$

This can also be written as a minimization problem of the negative log-likelihood:

$$\hat{\mathbf{x}} = \arg\min_{\substack{\mathbf{x} \\ |S|\leq K}} -\log p(\mathbf{y} \mid \mathbf{x}) = \arg\min_{\substack{\mathbf{x} \\ |S|\leq K}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2. \tag{6}$$

A crucial observation is that for any sparse vector $\mathbf{x}$, not all columns of the sensing matrix $\mathbf{A}$ are relevant. If we denote $\mathbf{x}_S$ as the vector containing only the non-zero entries of $\mathbf{x}$, indexed by the support S, and $\mathbf{A}_S$ as the submatrix of $\mathbf{A}$ containing only the columns corresponding to S, then the minimization problem becomes:

$$\hat{\mathbf{x}}_S = \underset{\substack{\mathbf{x}_S \\ |S| \leq K}}{\arg\min} \left\| \mathbf{y} - \mathbf{A}_S \mathbf{x}_S \right\|_2^2 . \tag{7}$$

This minimization problem has then split our problem in two steps. First we need to find the optimal support S; once it is known, the minimizer of the L2 norm is known to be the pseudo-inverse. In other words we get:

$$\hat{S} = \underset{\substack{S \\ |S| \leq K}}{\arg\min} \left\| \mathbf{y} - \mathbf{A}_S \mathbf{A}_S^+ \mathbf{y} \right\|_2^2 , \quad \hat{\mathbf{x}} = \mathbf{A}_{\hat{S}}^+ \mathbf{y}. \tag{8}$$

3 Putting it together

From equation (8) it can be seen that the CPWL function we learn indeed assigns linear transformations of the form $\mathbf{B} \in \mathbb{R}^{N \times M}$. Namely, it assigns the pseudo-inverse of the submatrix $\mathbf{A}_{\hat{S}}$ that minimizes the likelihood term. Thus, in order to get the number of unique matrices that the function assigns, we can count the number of unique supports S that can be assigned to the problem. Remember that the sparsity of the support is at most K and we have not shown that all supports are chosen (altough we do hypothesize this). Thus the optimal number of linear regions that can be assigned is bounded as:

$$\text{optimal number of projection regions} \leq \sum_{k=0}^{K} \binom{N}{k}. \tag{9}$$