This is exactly the same problem as the one presented in introduction\(^1\). This problem, which is identical to multi-stage gain-shape vector quantization [10], is illustrated in figure 2.

![Diagram](image)

**Fig. 2.** General scheme of the minimization problem.

Typical values for the different parameters greatly vary depending on the application. For example, in speech coding [20] (and especially for low bit rate) a highly redundant dictionary \((L \gg N)\) is used coupled with high sparsity \((K\) very small\)\(^2\). In music signals coding, it is common to consider much larger dictionaries and to select a much larger number of dictionary elements (or atoms). For example in the scheme proposed in [21] based on an union of MDCTs, the observed vector \(\underline{x}\) represents several seconds of the music signal sampled at 44.1 kHz and typical values could be \(N > 10^5\), \(L > 10^6\) and \(K \approx 10^3\).

### B. Standard iterative algorithm

If the indices \(j(1) \cdots j(K)\) are known (e.g. the matrix \(A\)), the solution is easily obtained following a least square minimization strategy [22]. Let \(\hat{x}\) be the best approximate of \(\underline{x}\), e.g. the orthogonal projection of \(\underline{x}\) on the subspace spanned by the column vectors of \(A\) verifying:

\[
< \underline{x} - Ag, f_{j(k)}> = 0 \text{ for } k = 1 \cdots K
\]

The solution is then given by

\[
g = (A^tA)^{-1}A^t\underline{x}
\]

\(^1\)Note though that the vector \(g\) is now of dimension \(K\) instead of \(L\). The indices \(j(1) \cdots j(K)\) point to dictionary vectors (columns of \(F\)) corresponding to non-zero gains.

\(^2\)\(K=2\) or 3, \(L = 512\) or 1024, \(N = 40\) for a sampling rate of 8kHz are typical values found in speech coding schemes.