Noise in Computation — Exercise 1 (submit June 26, 2007)

June 19, 2007

1. Fix a signal $f \in \mathbb{R}^N$. Let $x = f + w$ where w is zero mean white noise with variance σ^2. Let $B = (b_0, \ldots, b_{N-1})$ be an orthonormal basis for $\mathbb{R}^N$ and let $\hat{x}, \hat{f}, \hat{w}$ respectively denote the representation of x, f, w in the basis B. For $a = (a_0, \ldots, a_{N-1})$ let D_a be the following linear diagonal estimator for the signal f given observed signal x:

$$D_a(x) = \sum_{i=0}^{N-1} a_i \cdot \hat{x}_i \cdot b_i.$$

The *risk* of D_a with respect to f is defined as

$$r(D_a, f) = \mathbb{E}[||f - D_a(x)||^2],$$

where the expectation is taken over the random white noise w. (*i*) Find the "best possible" vector $a^* = (a_0^*, \ldots, a_{N-1}^*)$, i.e., such that $r(D_{a^*}, f)$ is minimal among all possible $a \in \mathbb{R}^N$. (*ii*) Calculate $r_{\inf}(f) = r(D_{a^*}, f)$.

2. Let $r_p(f)$ denote the minimal risk of a linear projection estimator of the signal f as defined in class:

$$r_p(f) = \sum_{i=0}^{N-1} \min(\hat{f}_i^2, \sigma^2).$$

Let D_T denote the threshold diagonal estimator given by

$$D_T(x) = \sum_{i=0}^{N-1} \rho_T(\hat{x}_i) \cdot b_i, \text{ where } \rho_T(\hat{x}_i) = \begin{cases} \hat{x}_i & |\hat{x}_i| \geq T \\ 0 & |\hat{x}_i| < T \end{cases}$$

Prove the following weaker form of the Donoho-Johnston Theorem presented in class:
Theorem: There exists N_0 such that for all $N > N_0$ and $f \in \mathbb{R}^N$ the following holds. For $T = \sigma \cdot (\sqrt{2 \ln N} + 1)$, with probability greater than 0.99

$$||f - D_T(x)||^2 \leq r_p(f) \cdot O(\ln N),$$

where $x = f + w$ and the probability is taken over w which is zero mean Guassian white noise. Hints:

- Represent all vectors in the basis B.

- Use the following Theorem.
  Theorem: There exists N_0 such that for all $N > N_0$

  $$\Pr\left[\sigma\sqrt{2\ln N} - \frac{\sigma \ln\ln N}{\ln N} \leq \max_{0\leq i<N} |w_i| \leq \sigma\sqrt{2\ln N}\right] > 0.99,$$

  where the probability is taken over independent identically distributed (i.i.d.) Gaussian random variables $w_0, \ldots, w_{N-1} \sim \mathcal{N}(0, \sigma^2)$ with zero mean and variance σ^2.
- Split the problem to several subcases, depending on $|\hat{f}_i|, \sigma$ and T.

3. Compare the risk of thresholding in different Basis. Let $r_p(f, B)$ denote the minimal risk of a linear projection estimator for f (as defined in question 2) when f is represented in the basis B. Let $N(f, B)$ denote the number of coefficients of the representation of f in basis B that have absolute value at least σ. In other words, if $\hat{f}$ is the representation of f in basis B, then

   $$N(f, B) = \left|\{\hat{f}_i \geq \sigma\}\right|.$$

   Let $C(f, B, m)$ be the error of reconstructing f from its m largest coefficients in basis B, i.e., assuming without loss of generality $\hat{f}_0 \geq \hat{f}_1 \geq \ldots \geq \hat{f}_{N-1}$,

   $$C(f, B, m) = \sum_{i>m} \hat{f}_i^2.$$

   (i) Express $r_p(f, B)$ in terms of $N(f, B)$, σ and $C(f, B, m)$. (Hint: What value should m receive?)

   (ii) Prove there exist two distinct orthonormal basis B_0, B_1 and two signals f_0, f_1 such that for $k \in 0, 1$ we have $r_p(f_k, B_k)$ is small while $r_p(f_{1-k}, B_k)$ is large?

   (iii) How large can the gap be?