Some statistical problems involving estimation of eigenvalues and eigenvectors from noisy data

October 6, 2006

1. **Functional linear regression :** Observe i.i.d. $\{(X_i, Y_i) : i = 1, \ldots, n\}$, where Y_i are scaler random variables and X_i are L^2 random processes on an interval $\mathcal{I}$. Assume the model:
$$Y_i = a + \int_{\mathcal{I}} b(u) X_i(u) du + \epsilon_i \tag{1}$$
where ϵ_i are i.i.d., zero mean r.v., $b \in L^2(\mathcal{I})$ is a deterministic function.

   - How to estimate b ?
   - How to predict Y ?

   One popular method is Principal Components Regression (PCR). (How many components to select ?)

2. **Factor model in statistics :** Observe i.i.d. data X_i distributed as
$$X_i = \mu + \sum_{k=1}^{M} \sqrt{\lambda_k} v_{ki} \theta_k + \sigma Z_i, \tag{2}$$
where $\{v_{ki} : 1 \le k \le M, 1 \le i \le n\}$ are i.i.d. $N(0,1)$ r.v. and $\{Z_i\}$ are i.i.d. $N(0, I)$. $\theta_k^T \theta_l = \delta_{kl}$ for $1 \le k, l \le M$.

   - How to estimate M, $\{(\lambda_k, \theta_k) : k = 1, \ldots, M\}$?
   - What if data consist of irregularly sampled functions from an interval ?
   - More general covariance matrix for Z ?

3. **Factor model in finance :** A market involving K companies, and B industrial branches. κ_b companies in b-th industrial branch ($b = 1, \ldots, B$) and $\sum_{b=1}^{B} \kappa_b = K - kappa$, where $0 \le \kappa \le K$ is the number of companies not associated with any industrial branch. If company k is in branch b, then series of its returns: $M_k(t)$ represented as:
$$M_k(t) = q_b \eta_b(t) + r_b \epsilon_k(t), \tag{3}$$
where $q_b, r_b > 0$ and $\eta_b(t)$,$\epsilon_k(t)$ are i.i.d. $N(0, 1)$.

4. **Time dependent observations :**

   - Vector autoregressive process.
   - Autoregressive Hilbertian process.