Supplemental file for the paper "Discovering the Nature of Variation in Nonlinear Profile Data" by Zhenyu Shi, Daniel W. Apley and George C. Runger

Dose-Response Profile Example

This supplement illustrates the manifold learning approach applied to the dose-response profile example introduced in Section 5 of the paper. The objective is to discover the nature of the profile-to-profile variation and to parameterize it via some low dimensional representation with parameters that are in agreement with the randomly varying parameters $\boldsymbol{\theta}_i$ in the model (1). In the following, we randomly generate a sample of profiles from the model (1) with $\boldsymbol{\theta}_i$ varying randomly from profile-to-profile. After generating the profiles we subsequently treat the model (1) as unknown, and we use manifold learning to discover the nature of the profile-to-profile variation. To be successful at this objective, the estimated manifold $\hat{\mathbf{f}}(\mathbf{v})$ and its manifold coordinate representation $\{\hat{\mathbf{v}}_i\colon\ i = 1\,, 2\,, 3\,, \ldots, N\}$ over the sample should be in agreement with the actual manifold $\mathbf{f}(\boldsymbol{\theta})$ and its parametric representation $\{\boldsymbol{\theta}_i\colon\ i = 1\,, 2\,, 3\,, \ldots, N\}$.

To generate the sample of profiles, we generated N = 200 random realizations of θ_{3i} and θ_{4i}, and for each realization $\{\theta_{3i}, \theta_{4i}\}$ we used the model (1) to generate a profile at n = 30 evenly spaced dosages over the interval [1, 30] with $\{\theta_{1i} = 0, \theta_{2i} = 1\}$ held fixed and $w_{ji} \sim NID(0, \sigma^2)$ with noise standard deviation $\sigma = 0.04$. The sample of profiles are plotted in Figure 1. Because only two parameters were varied, the profiles lie (aside from noise) on a manifold of dimension p = 2 in n-dimensional space. Figure 2 is a plot of the first three PCA scores, which shows a clear nonlinearity in the manifold.

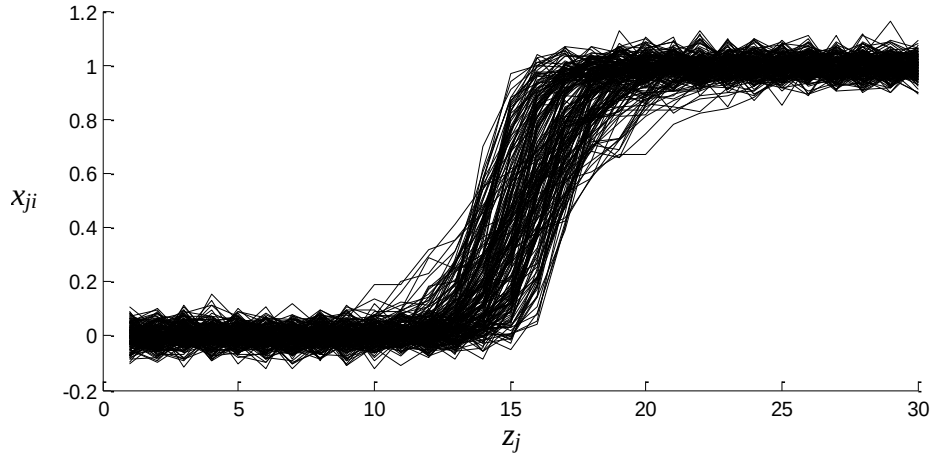

Figure 1. Sample of 200 dose-response profiles generated from the model (1) with $\{\theta_{3i}, \theta_{4i}\}$ varied randomly.

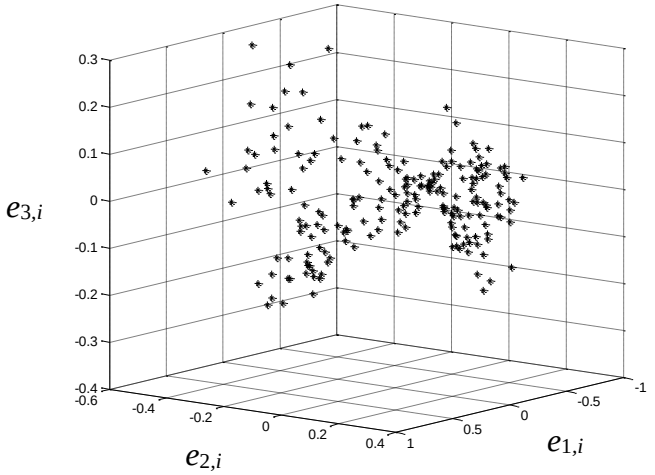

Figure 2. Scatter plot of the first three PCA scores for the 200 dose-response profiles in Figure 1.

Figure 3 shows the estimated manifold coordinates $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^{\mathrm{T}}: i = 1, 2, 3, \ldots, 200\}$ from the ISOMAP manifold learning algorithm applied to the first P = 8 PCA scores. Figure 4 illustrates the estimated manifold $\hat{\mathbf{f}}(\mathbf{v})$, estimated via a multi-response neural network, in the profile geometry space (analogous to Figure 11(b) in the paper). Figure 4(a) illustrates the nature of the first variation source v_1, which corresponds closely to the effects of θ_3 in (1); and Figure 4(b) illustrates the nature of the second variation source v_2, which corresponds closely to the effects of θ_4 in (1). In this respect, the manifold learning and visualization approach has effectively discovered the nature of the two major sources of variation in the dose-response profiles.

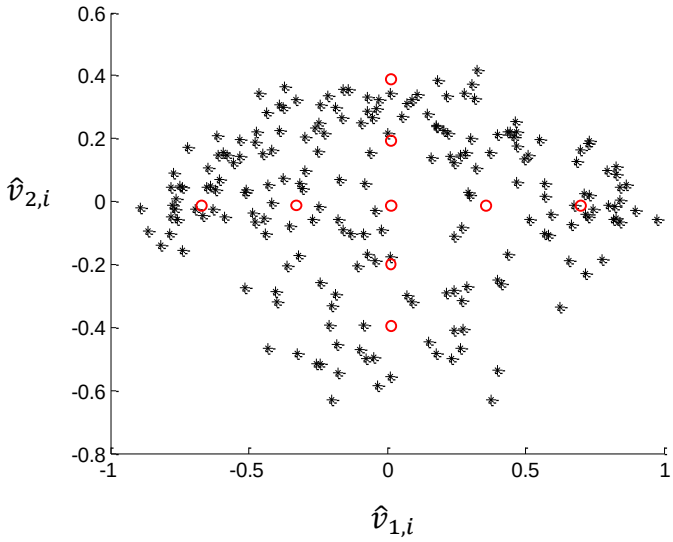

Figure 3. Estimated manifold coordinates $\{\hat{\mathbf{v}}_i = [\hat{v}_{1,i}, \hat{v}_{2,i}]^{\mathrm{T}}: \ i = 1, 2, 3, \ldots, 200\}$ for the dose-response profiles in Figure 1.

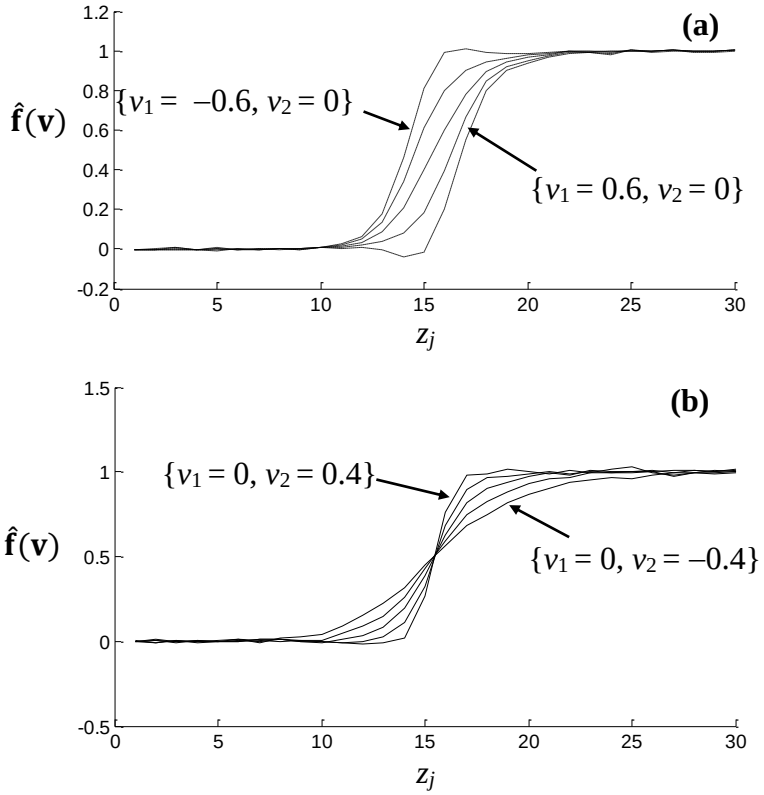

Figure 4. Illustration of the estimated manifold $\hat{\mathbf{f}}(\mathbf{v})$ in the profile geometry space for the dose-response profiles in Figure 1. Panel (a) illustrates the effects of v_1 by plotting $\hat{\mathbf{f}}(\mathbf{v})$ for the five values of $\mathbf{v}$ corresponding to the horizontal locus of open red circles in Figure 3. Panel (b) illustrates the effects of v_2 by plotting $\hat{\mathbf{f}}(\mathbf{v})$ for the five values of $\mathbf{v}$ corresponding to the vertical locus of open red circles in Figure 3.