1.1.5 A related open problem

We now show an interesting open problem posed by Mallat and Zeitouni at [MZ11]

Open Problem 1.1 (Mallat and Zeitouni [MZ11]) *Let $g \sim \mathcal{N}(0, \Sigma)$ be a gaussian random vector in $\mathbb{R}^p$ with a known covariance matrix Σ and $d < p$. Now, for any orthonormal basis $V = [v_1, \ldots, v_p]$ of $\mathbb{R}^p$, consider the following random variable Γ_V: Given a draw of the random vector g, Γ_V is the squared ℓ_2 norm of the largest projection of g on a subspace generated by d elements of the basis V. The question is:*

What is the basis V for which $\mathbb{E}\left[\Gamma_V\right]$ is maximized?

The conjecture in [MZ11] is that the optimal basis is the eigendecomposition of Σ. It is known that this is the case for $d = 1$ (see [MZ11]) but the question remains open for $d > 1$. It is not very difficult to see that one can assume, without loss of generality, that Σ is diagonal.

A particularly intuitive way of stating the problem is:

1. Given $\Sigma \in \mathbb{R}^{p \times p}$ and d
2. Pick an orthonormal basis $v_1, \ldots, v_p$
3. Given $g \sim \mathcal{N}(0, \Sigma)$
4. Pick d elements $\tilde{v}_1, \ldots, \tilde{v}_d$ of the basis
5. `Score`: $\sum_{i=1}^{d} \left(\tilde{v}_i^T g\right)^2$

The objective is to pick the basis in order to maximize the expected value of the `Score`.

Notice that if the steps of the procedure were taken in a slightly different order on which step 4 would take place before having access to the draw of g (step 3) then the best basis is indeed the eigenbasis of Σ and the best subset of the basis is simply the leading eigenvectors (notice the resemblance with PCA, as described above).

More formally, we can write the problem as finding

$$\underset{\substack{V \in \mathbb{R}^{p \times p} \\ V^T V = I}}{\operatorname{argmax}} \left(\mathbb{E}\left[\max_{\substack{S \subset [p] \\ |S| = d}} \sum_{i \in S} \left(v_i^T g\right)^2 \right] \right),$$

where $g \sim \mathcal{N}(0, \Sigma)$. The observation regarding the different ordering of the steps amounts to saying that the eigenbasis of Σ is the optimal solution for

$$\underset{\substack{V \in \mathbb{R}^{p \times p} \\ V^T V = I}}{\operatorname{argmax}} \left(\max_{\substack{S \subset [p] \\ |S| = d}} \mathbb{E}\left[\sum_{i \in S} \left(v_i^T g\right)^2 \right] \right).$$

Reference

[MZ11] S. Mallat and O. Zeitouni. A conjecture concerning optimality of the karhunen-loeve basis in nonlinear reconstruction. *Available online at arXiv:1109.0489 [math.PR]*, 2011.