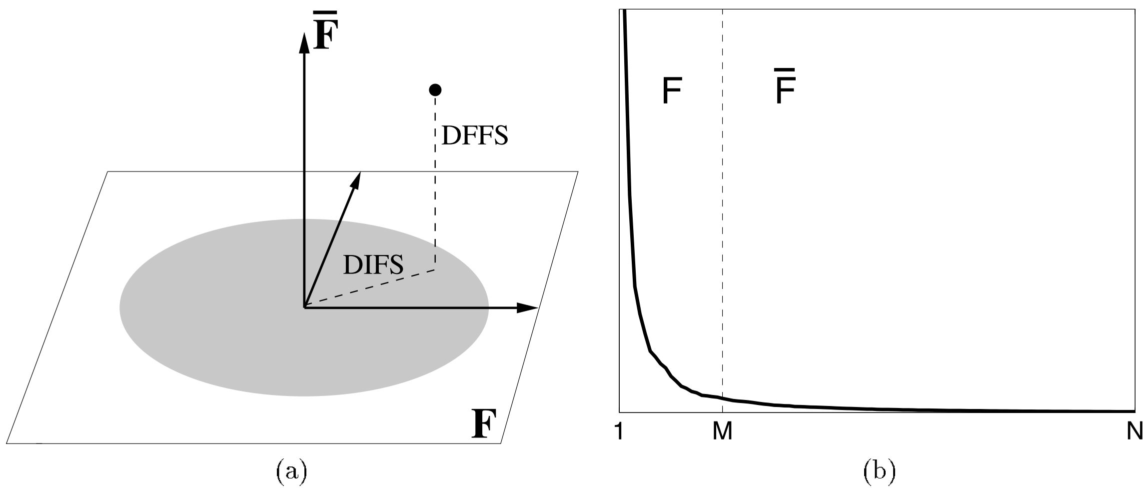

Figure 1: (a) Decomposition of $\mathcal{R}^N$ into the principal subspace F and its orthogonal complement $\bar{F}$ for a Gaussian density, (b) a typical eigenvalue spectrum and its division into the two orthogonal subspaces.

value for the weighting parameter ρ — found by minimizing cross-entropy — is simply the average of the $\bar{F}$ eigenvalues[1]

$$\rho = \frac{1}{N-M} \sum_{i=M+1}^{N} \lambda_i \tag{5}$$

We note that in actual practice, the majority of the $\bar{F}$ eigenvalues are unknown but *can* be estimated, for example, by fitting a nonlinear function to the available portion of the eigenvalue spectrum and estimating the average of the eigenvalues beyond the principal subspace.

3.2 Efficient Similarity Computation

Consider a feature space of Δ vectors, the differences between two images (I_j and I_k). The two classes of interest in this space correspond to intrapersonal and extrapersonal variations and each is modeled as a high-dimensional Gaussian density

$$\begin{aligned} P(\Delta|\Omega_E) &= \frac{e^{-\frac{1}{2}\Delta^T \Sigma_E^{-1} \Delta}}{(2\pi)^{D/2}|\Sigma_E|^{1/2}} \\ P(\Delta|\Omega_I) &= \frac{e^{-\frac{1}{2}\Delta^T \Sigma_I^{-1} \Delta}}{(2\pi)^{D/2}|\Sigma_I|^{1/2}} \end{aligned} \tag{6}$$

The densities are zero-mean since for each $\Delta = I_j - I_k$ there exists a $\Delta = I_k - I_j$. Since these distributions are known to occupy a principal subspace of image space (face-space), only the principal eigenvectors of the Gaussian densities are relevant for modeling. These densities are used to evaluate the *similarity score* in Eq. 2 in accordance with the density estimate in Eq. 4.

Computing the similarity score involves first subtracting a candidate image I_j from a database entry I_k. The resulting Δ is then projected onto the principal eigenvectors of both extrapersonal and intrapersonal Gaussians. The exponentials are then evaluated, normalized and combined as likelihoods in Eq. 2. This operation is iterated over all members of the database (many I_k images) until the maximum score is found (i.e. the match). Thus, for large databases, this evaluation is rather expensive.

[1] Tipping & Bishop [33] have since derived the same estimator for ρ by showing that it's a saddle point of the likelihood for a latent variable model.