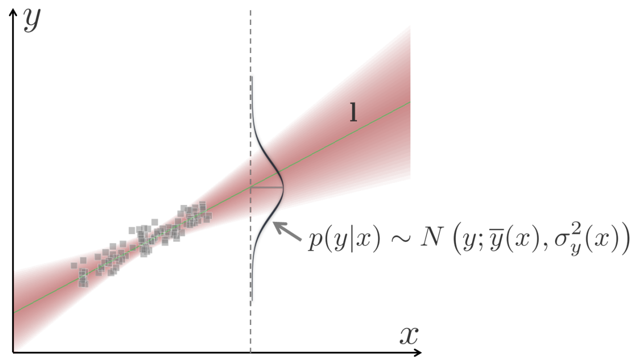

Fig. 3: **Probabilistic line fitting.** Given a set of training points we can fit a line model to them. For instance, in this example $\mathbf{l} \in \mathbb{R}^2$. Matrix perturbation theory enables us to compute the entire conditional density $p(\mathbf{l}|x)$ from where we can derive $p(y|x)$. Training a regression tree involves minimizing the uncertainty of the prediction $p(y|x)$. Therefore, the training objective is a function of σ_y^2.

Finally the 3×3 line covariance matrix is

$$\Lambda_{\mathbf{l}} = \mathtt{J}\ \mathtt{S}\ \mathtt{J} \tag{1}$$

with the 3×3 Jacobian matrix

$$\mathtt{J} = -\sum_{k=2}^{3} \frac{\mathbf{u}_k \mathbf{u}_k^\top}{\lambda_k}$$

where λ_k denotes the k^{th} eigenvalues of the matrix $\mathtt{M}$ and $\mathbf{u}_k$ its corresponding eigenvector. The 3×3 matrix $\mathtt{S}$ in (1) is

$$\mathtt{S} = \sum_{i=1}^{n} \left(\mathbf{a}_i^\top \mathbf{a}_i \mathbf{l}^\top \Lambda_i \mathbf{l} \right).$$

Therefore the distribution over $\mathbf{l}$ remains completely defined. Now, given a set of (x, y) pairs we have found the maximum-likelihood line model $\mathcal{N}\left(\bar{\mathbf{l}}, \Lambda_l\right)$. However, what we want is the conditional distribution $p(y|x)$ (see fig. 3) this is discussed next.