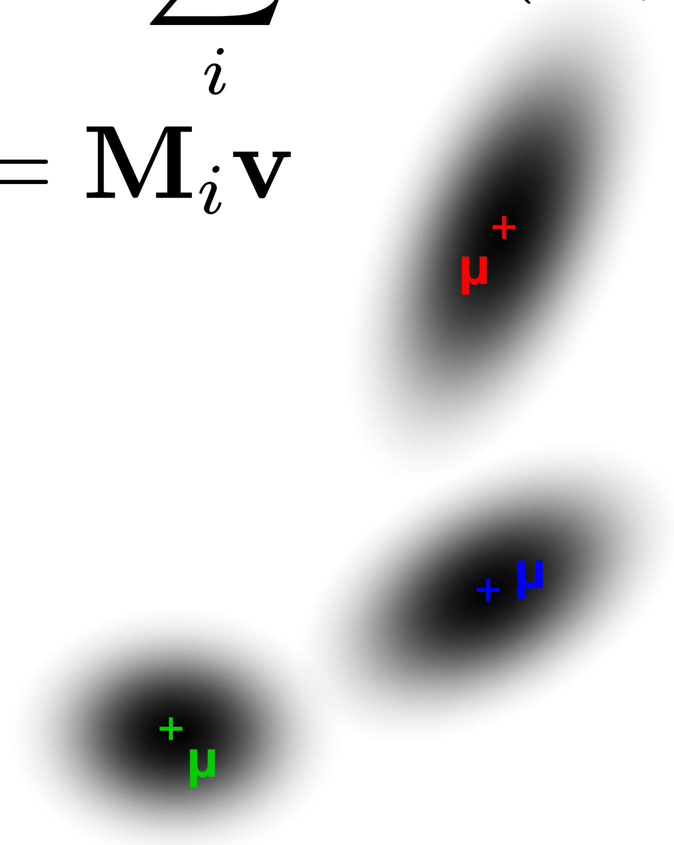
Simple Example of Subspace GMM Model

Subspace GMM Model Example

$$p(\mathbf{x}) = \sum_i w_i \mathcal{N}(\mathbf{x}; \mu_i, \mathbf{\Sigma}_i)$$

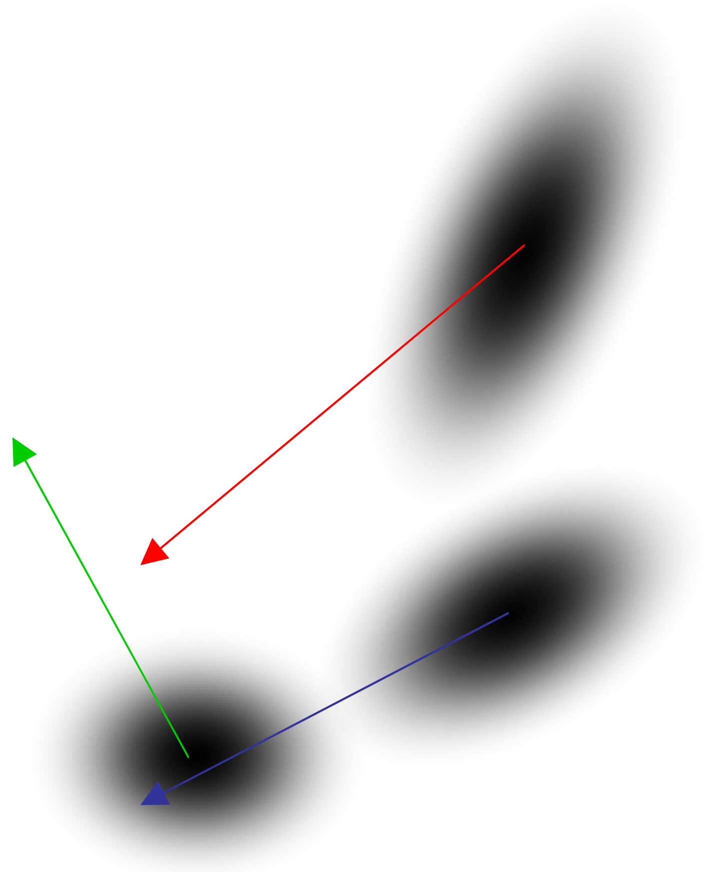
$$\mu_i = \mathbf{M}_i \mathbf{v}$$



$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

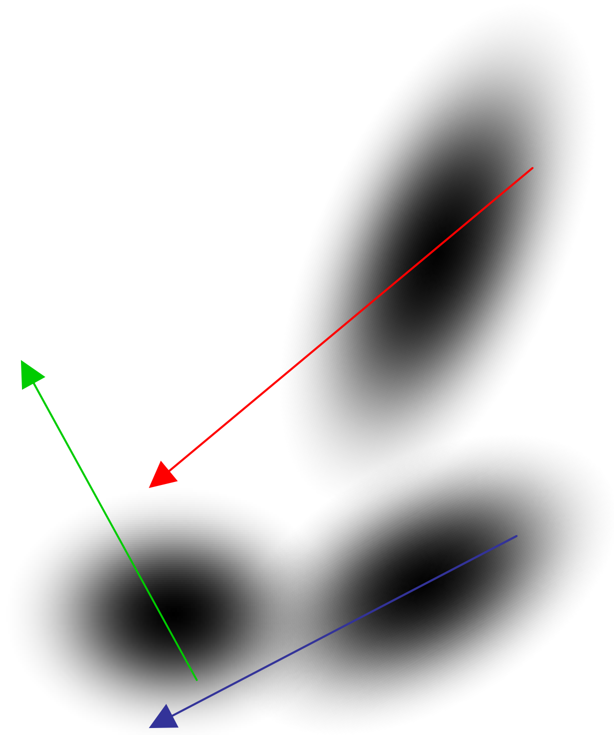
Let w_i and $\sum_i$ be fixed in our model for now.

Subspace GMM Model Example



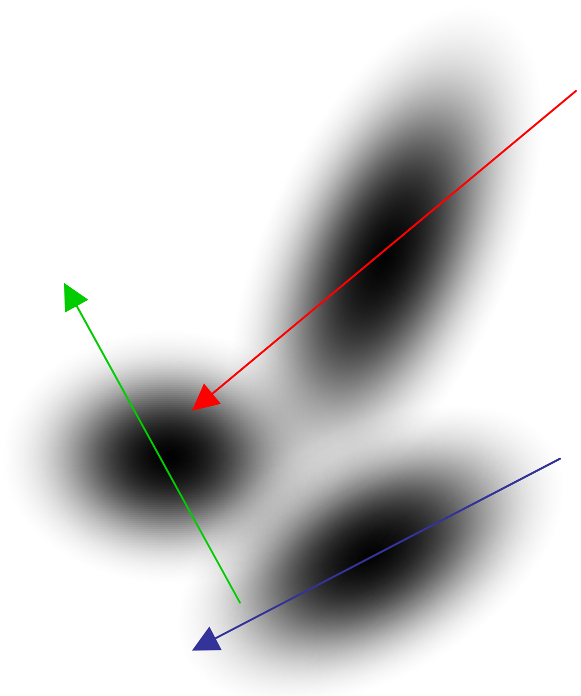
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



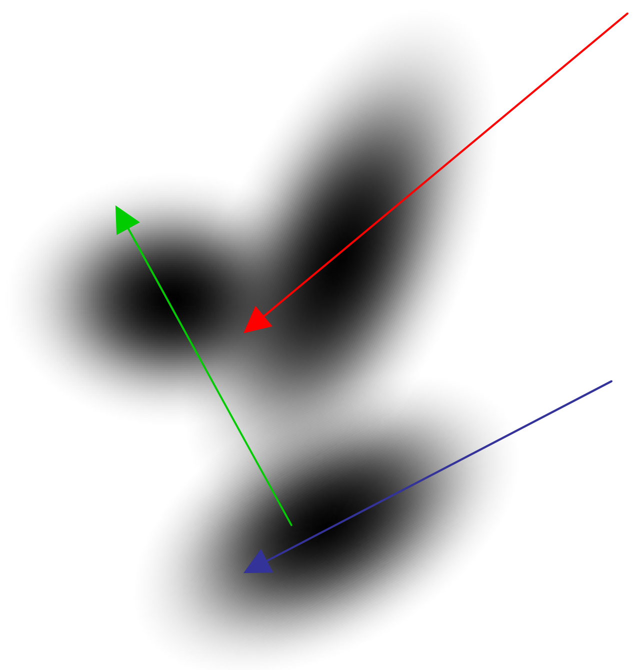
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



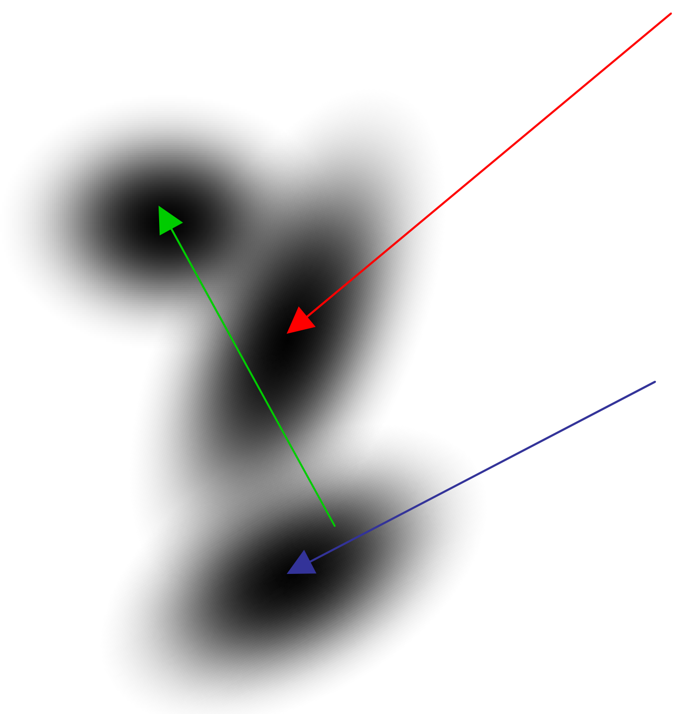
$$\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_1\\ \mu_2\\ \mu_1\\ \mu_2\end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13}\\ m_{21} & m_{22} & m_{23}\\ m_{11} & m_{12} & m_{13}\\ m_{21} & m_{22} & m_{23}\\ m_{11} & m_{12} & m_{13}\\ m_{21} & m_{22} & m_{23}\end{bmatrix} \begin{bmatrix}\mathbf{v_1}\\ v_2\\ v_3\end{bmatrix}$$

Subspace GMM Model Example



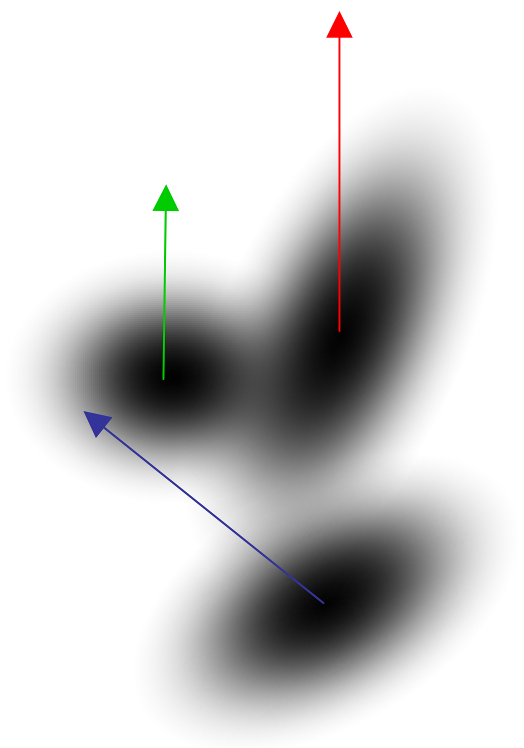
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



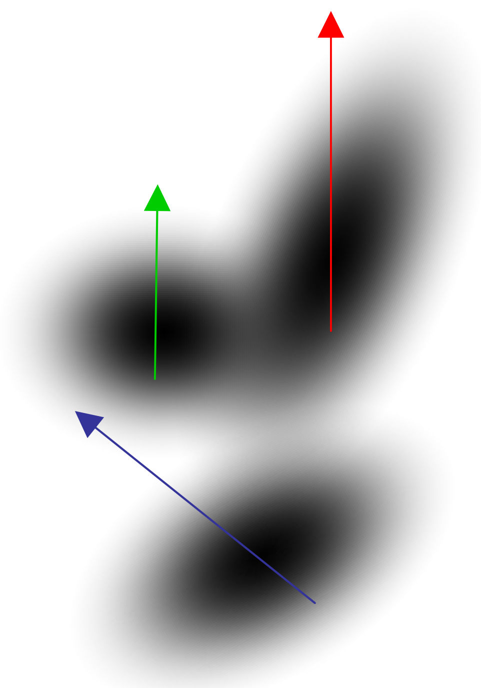
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



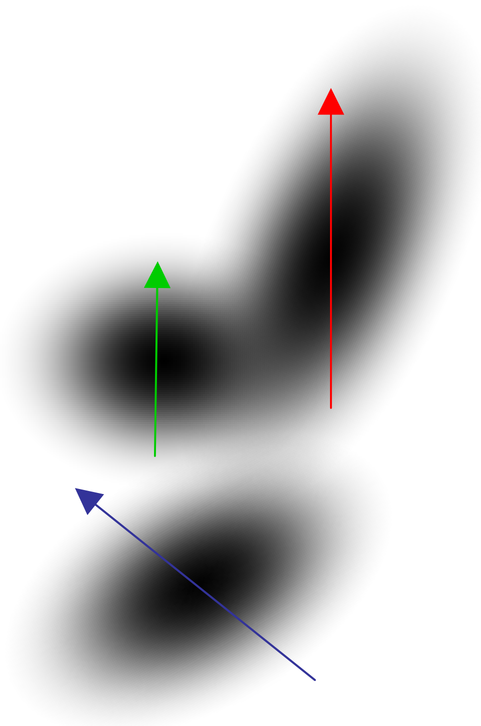
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



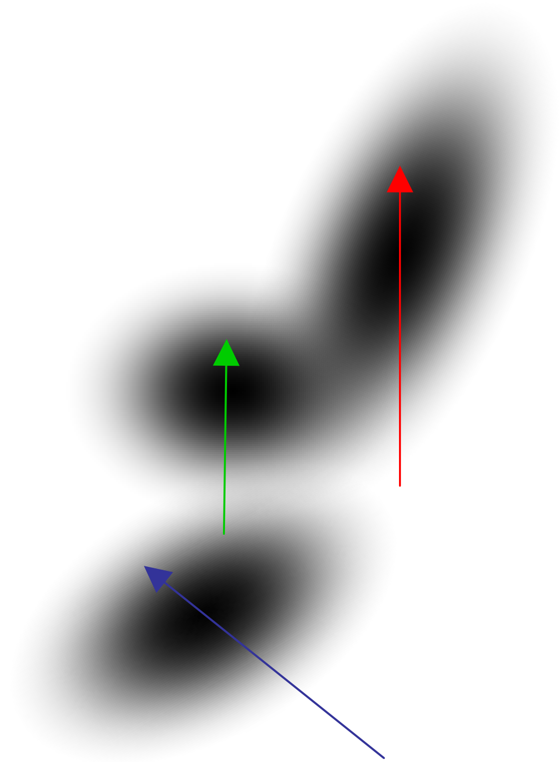
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



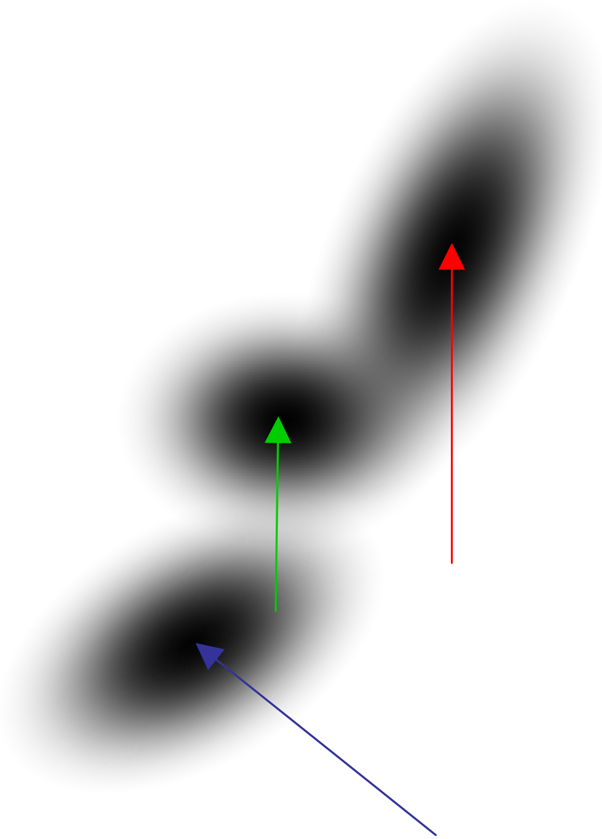
$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \\ v_3 \end{bmatrix}$$

Subspace GMM Model Example



$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ v_3 \end{bmatrix}$$

Modeling mixture weights

- Log-linear model is used for modeling mixture weights

$$
\mathrm{w}_i = \frac{\exp \mathbf{w}_i^T \mathbf{v}}{\sum_k \exp \mathbf{w}_k^T \mathbf{v}}
$$

$$
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \overline{\mathrm{w}} \\ \log \overline{\mathrm{w}} \\ \log \overline{\mathrm{w}} \end{bmatrix}
=
\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix}
\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}
$$

Modeling mixture weights

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \bar{w} \\ \log \bar{w} \\ \log \bar{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Modeling mixture weights

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \bar{w} \\ \log \bar{w} \\ \log \bar{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

Modeling mixture weights

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \overline{w} \\ \log \overline{w} \\ \log \overline{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

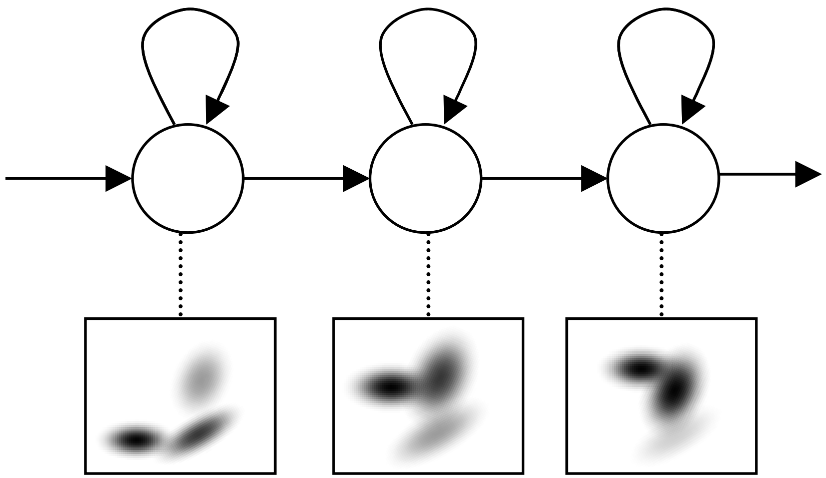
Modeling mixture weights

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \overline{w} \\ \log \overline{w} \\ \log \overline{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

Modeling mixture weights

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log \overline{w} \\ \log \overline{w} \\ \log \overline{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ v_2 \\ v_3 \end{bmatrix}$$

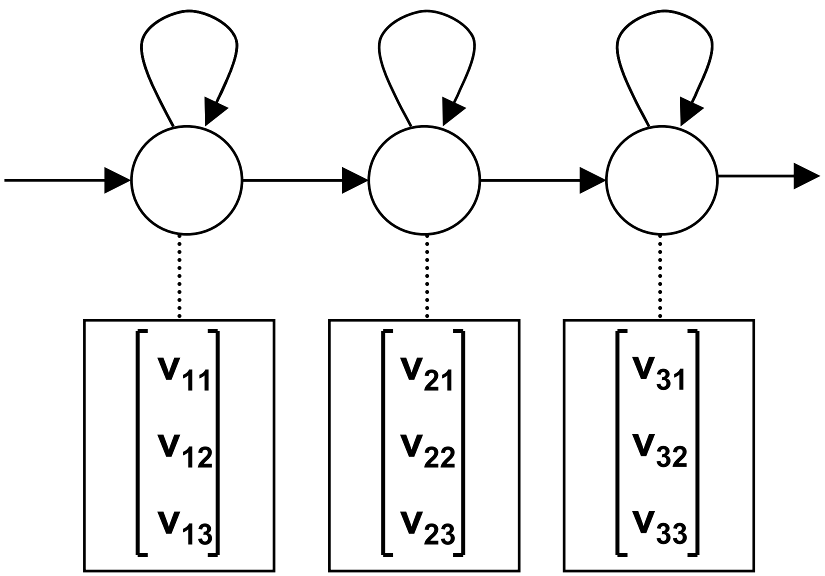
Acoustic model for speech recognition



- Speech sounds are typically modeled by HMMs with state distributions given by GMMs.
- Typically, there are thousands of such models corresponding to context dependent phonemes.
- Many state distributions are very similar and exhibit certain regularities.

Acoustic Model with Subspace GMM

Parameters shared across HMM states (includes also covariance matrices)



$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log w \\ \log w \\ \log w \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

State specific parameters are low dimensional vector

Controlling ratio between shared and state specific parameters

- Increasing number of Gaussian component increase number of shared parameters
- Increasing size of vector **v** increase number of both shared and state specific parameters
- It would be useful to have the possibility of increasing number of state specific parameters without affecting the number of shared parameters

$$
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \log w \\ \log w \\ \log w \\ \log w \end{bmatrix}
=
\begin{bmatrix}
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
w_1 & w_2 & w_3 & w_4 \\
w_1 & w_2 & w_3 & w_4 \\
w_1 & w_2 & w_3 & w_4 \\
w_1 & w_2 & w_3 & w_4
\end{bmatrix}
\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}
$$

Substates – mixture of subspace GMM distributions

- In our experiments, we keep splitting substates to reach the best performance
- Can be seen as an alternative to splitting Gaussians in standard HMM system

Mixture weights

$$
\begin{bmatrix}
\mu_1 & \mu_1 \\
\mu_2 & \mu_2 \\
\mu_1 & \mu_1 \\
\mu_2 & \mu_2 \\
\mu_1 & \mu_1 \\
\mu_2 & \mu_2 \\
\log w & \log w \\
\log w & \log w \\
\log w & \log w
\end{bmatrix}
=
\begin{bmatrix}
m_{11} & m_{12} & m_{13} \\
m_{21} & m_{22} & m_{23} \\
m_{11} & m_{12} & m_{13} \\
m_{21} & m_{22} & m_{23} \\
m_{11} & m_{12} & m_{13} \\
m_{21} & m_{22} & m_{23} \\
w_1 & w_2 & w_3 \\
w_1 & w_2 & w_3 \\
w_1 & w_2 & w_3
\end{bmatrix}
\begin{array}{c}
\begin{array}{cc} c & c \end{array} \\
\begin{bmatrix}
v_1 & v_1 \\
v_2 & v_2 \\
v_3 & v_3
\end{bmatrix}
\end{array}
$$

Complete Model Definition (so far)

Parameters shared across HMM states	State specific parameters

Gaussian

HMM state

Substate

$$\mu_{jmi} = \mathbf{M}_i \mathbf{v}_{jm}$$

$$\mathrm{w}_{jmi} = \frac{\exp \mathbf{w}_i^T \mathbf{v}_{jm}}{\sum_h \exp \mathbf{w}_h^T \mathbf{v}_{jm}}$$

$$p(\mathbf{x}|j) = \sum_m c_{jm} \sum_i w_{jmi} \mathcal{N}(\mathbf{x}; \mu_{jmi}, \mathbf{\Sigma}_i)$$

Experimental part

Overview

- Baseline system
- Subspace system results
- Multilingual setup and results
- Training on very limited amount of data
- Interpreting subspace dimensions

Baseline - data

Acoustic data: CallHome databases

Language	Training set length	Evaluation set length
English	15.1h	1.8h
Spanish	16.5h	2.0h
German	14.7h	3.7h

Language model training:

- English: CallHome, Switchboard I, Switchboard Cellular, GigaWord and web data
- Spanish: CallHome and web data

Baseline systems

- PLP features
- Unadapted ML trained triphone models
- 16 Gaussians per state
- Bi-gram LM for English, tri-gram LM for Spanish
- No LVCSR build for German; results will be reported in terms of phone recognition performance
- The results are in agreement with those reported by other sites on this challenging task

	Accuracy (%)
CallHome English	45.3
CallHome Spanish	31.1

English subspace model training

- Initial configuration:
  - 1921 states
  - 400 Gaussians components
  - 39 dimensional features
  - 40 dimensional state vector – $\mathbf{v}_{jkm}$
  - **952k shared parameters**
  - **77k state specific parameters** (for single substate per state)
- Initial state alignment is taken from baseline system, later realigned by the model itself

Initial results for English

	Shared parameters	State-specific parameters	Accuracy (%)
Baseline	0	2427k	45.3
SGMM, 2k substates	952k	77k	47.5
SGMM, 9k substates	952k	363k	50.3

- For SGMM model, the number of state specific parameters is only a fraction of the number of shared parameters

Initial results for English

	Shared parameters	State-specific parameters	Accuracy (%)
Baseline	0	2427k	45.3
SGMM, 2k substates	952k	77k	47.5
SGMM, 9k substates	952k	363k	50.3

- Increasing the number of substates allow us to balance the ratio between the state specific and the shared parameters
- Still the overall number of the parameters in the SGMM model is less than half compared to the baseline

Searching for optimal configuration

- Tunable parameters:
  - number of Gaussian
  - number of tied states
  - number of substates
  - state vector dimension

- We did not find SGMM to be sensitive to exact setting of the parameters

- Best configuration found was with 3937 tied states, 16k substates, 400 Gaussians and state vector dimension 40

Accuracy = 50.8 %

Multilingual experiments

- Can data from another languages help to estimate share parameters more precisely?
- English, Spanish and German recognizers are trained together, where
  - each language has its own state specific parameters
  - shared parameters are shared also across languages
  - shared parameters are now trained on 46.3h of training data (English: 15.1h, Spanish: 16.5h, German: 14.7h)

Word recognition experiments

- English system

System	Shared parameters	State-specific parameters	Accuracy (%)
baseline	0	2427k	45.3
English only, 400 G	952k	363k	50.3
All languages, 800 G	1904k	890k	52.1

- Spanish system

System	Shared parameters	State-specific parameters	Accuracy (%)
baseline	0	2006k	31.1
Spanish only, 400 G	952k	312k	34.8
All languages, 800 G	1904k	762k	36.0

Phoneme recognition experiments

- Bigram phonotactic language models were trained on CallHome training sets
- Phoneme recognition accuracy is evaluated

System / Language	English	Spanish	German
# phonemes	42	27	45
baseline	45.1	53.8	43.9
Language only, 400 G	48.3	56.0	46.6
All languages, 800 G	49.8	56.3	47.4

- Training shared parameters across languages results in improved recognition performance for all the languages
- We benefit from increasing the number of shared parameters, which are now trained on more data

Experiments with limited amount of training data

- Can subspace model help us to build recognizer for a language with very limited amount of training data?
- English recognizers is trained, where
  - Shared parameters are trained on Spanish (16.5h) and German (14.7h) data
  - state specific parameters are trained on 1 hour of English

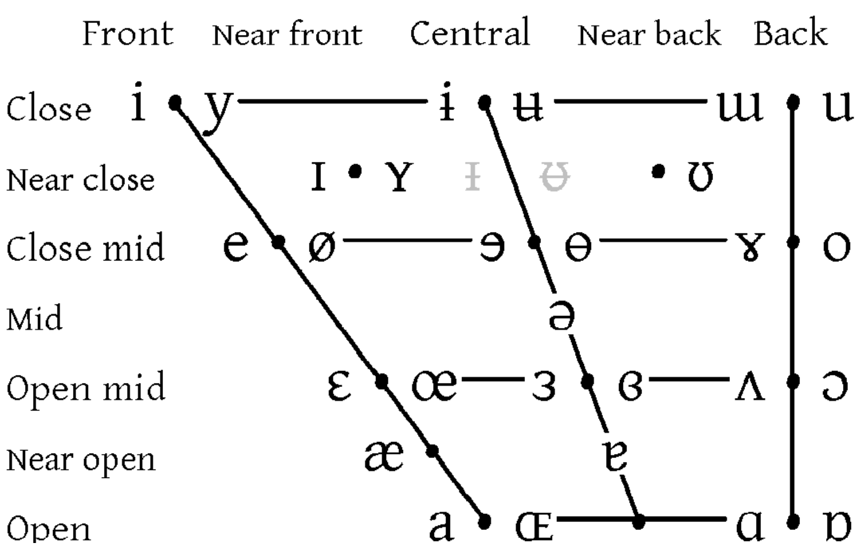
1 hour of training data

System	Accuracy (%)
HTK system, 500 tied states	27.6
SGMM, 1000 tied states, 20 dim, trained on English only	30.9
SGMM, 1500 tied states, 40 dim, shared parameters trained on Spanish + German	37.6

Interpreting subspace dimensions

- The state specific vectors $\mathbf{v}_{jkm}$ are relatively low-dimensional. Can we make the dimension even lower and visualize them?
- Substate system with 5 dimensions was trained
  - the accuracy is 34.2%
  - two most significant dimensions are shown

Vowel chart



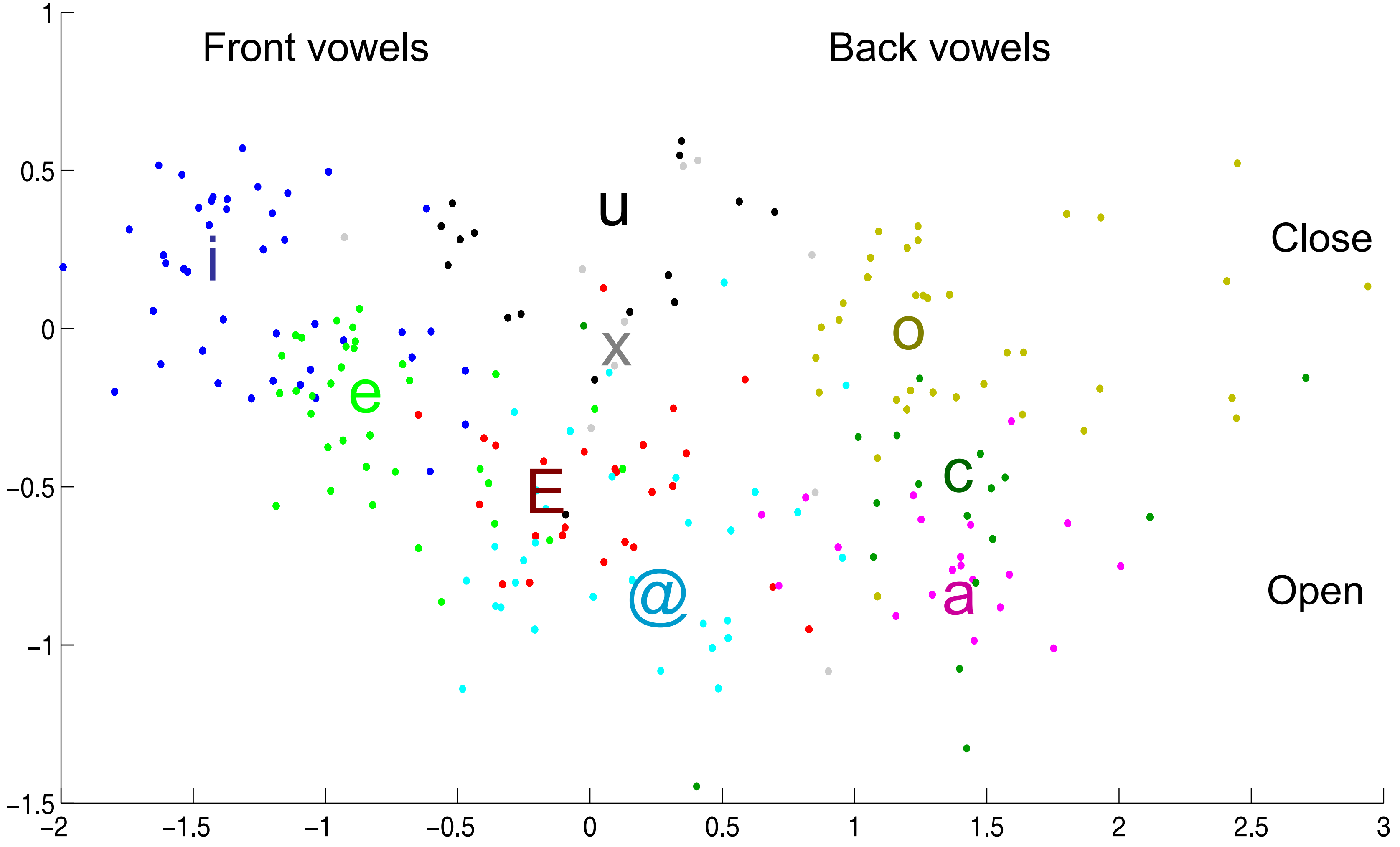
i u

e x o

E c

@ a

Phoneme (state) space



Conclusions

- Subspace GMM system outperform classical GMM system
- Training of subspace GMM shared parameters on multiple languages gives us an advantage
- Subspace GMM system can be successfully used for very limited amount of training data
- Subspace GMM system allows us to visualize state specific parameters. This gives us insight to the system and can serve as an analysis tools.