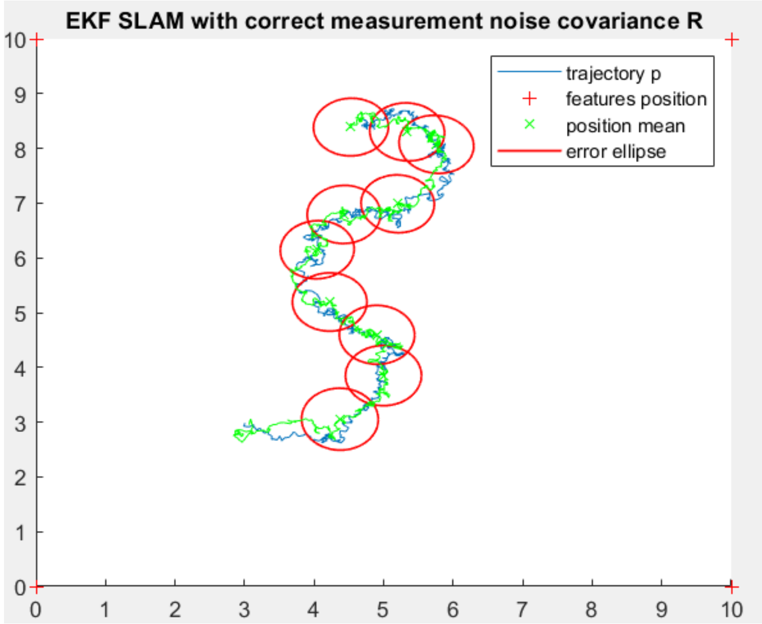
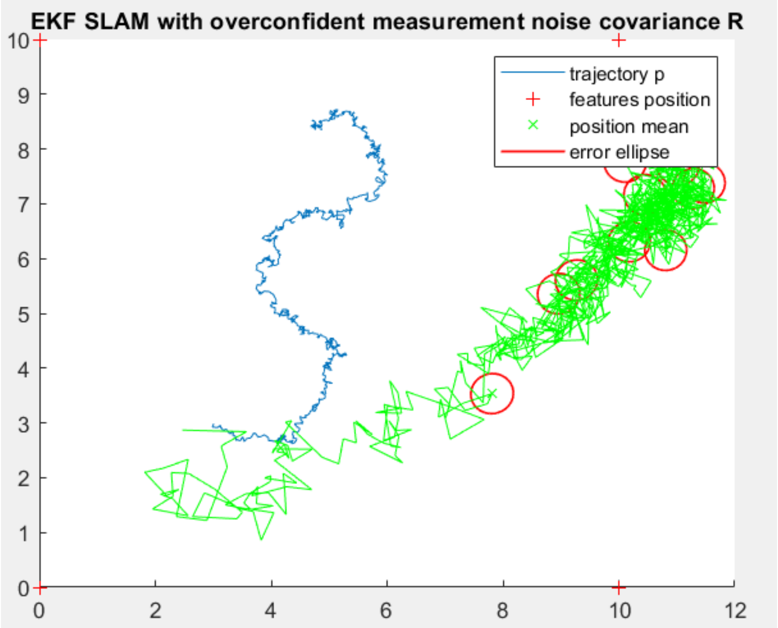
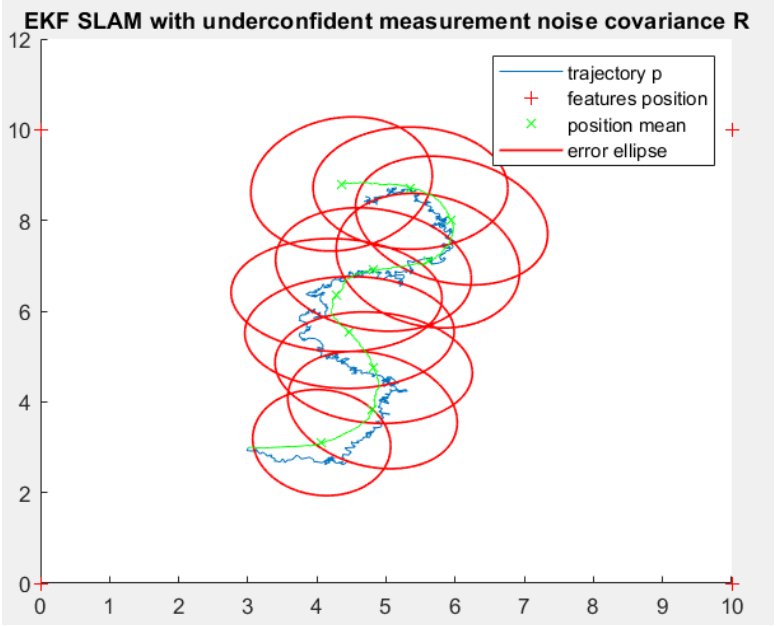
The Issue of Unknown Measurement Noise

In class: for EKF, we assumed that we knew the measurement covariance matrix R.

What happens if we don't?







Measurement Noise	Correct estimate	Overconfident	Underconfident
Noise Covariance	$R = I_8$	$R_{over} = 0.01R$	$R_{under} = 100R$
Average L2-distance	0.0077	0.0336	0.0133
Proportion of DV	11%	84%	2%

Naive Approach: Online Empirical Estimation

Idea: correct your measurement noise prior with incoming measurements

Updated EKF algorithm

- Predict step:

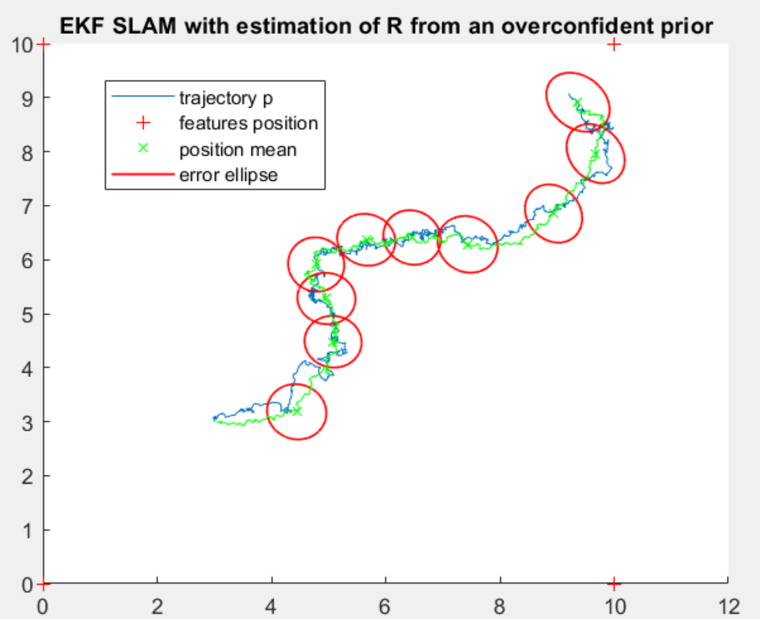
$\mu_{t|t-1}\,;\,\Sigma_{t|t-1}$

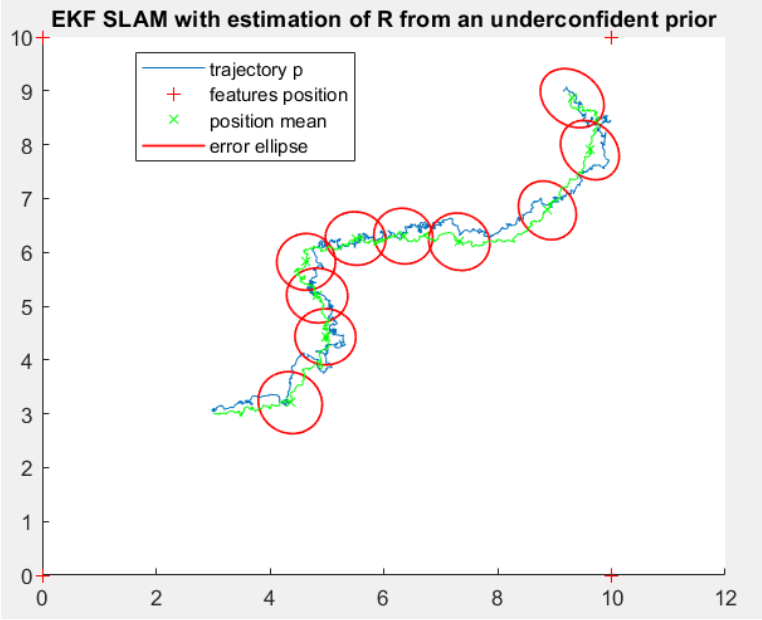
- Noise estimation: $\rho = 0.9$

$$R_t = \rho\, R_{t-1} + (1-\rho)\left(y_t - h(\mu_{t|t-1})\right)_{outer}$$

- Update step:

$\mu_{t|t}\,;\,\Sigma_{t|t}$





Measurement Noise	Correct estimate	Overconfident Prior	Underconfident Prior
Noise Covariance	$R = I_8$	$R_{over,0} = 0.01R$	$R_{under,0} = 100R$
Average L2-distance	0.0074	0.0065	0.0067
Proportion of DV	11%	5%	5%

Distributed Filtering using Alternating Direction Method of Multipliers (ADMM)

Gael Colas Michael Gobble

- All Bayesian Filter formulations for linear Gaussian processes can be formulated as optimizing a sum of quadratic functions
- The simplest MAP estimation step is a single Kalman Filter update of the form

$$\mu_{t|t} = \underset{x_{t|t}}{argmin} \; \frac{1}{2}\Big(y_{t|t} - C_t x_{t|t} - D_t u_t)\Big)^T R^{-1}\Big(y_{t|t} - C_t x_{t|t} - D_t u_t)\Big) \;+\; \frac{1}{2}\Big(x_{t|t} - A_t \mu_{t-1|t-1} - B_t u_t)\Big)^T \Big(A\Sigma_{t-1|t-1}A^T + Q\Big)\Big(x_{t|t} - A_t \mu_{t-1|t-1} - B_t u_t)\Big)$$

- If R is large and dense, this cannot be easily solved
- If R is diagonal or sparse, this can be solved very efficiently

Distributed Filtering using Alternating Direction Method of Multipliers (ADMM)

Gael Colas   Michael Gobble

- Goal 1: Compute a sparse approximation of R^{-1}
- Goal 2: Decompose objective into a large sum of simple quadratic functions and solve each in parallel

- Method: Use ADMM algorithm to achieve both goals

Alternating Direction Method of Multipliers

- An iterative algorithm for solving convex optimization problems in parallel
- Given an objective function of the form

$$\text{minimize} \quad f(x) = \textstyle\sum_{i=1}^{N} f_i(x)$$

where each function can be minimized separately in parallel

- The ADMM implementation runs the following algorithm to convergence

$$\begin{aligned} x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} \left(f_i(x_i) + y_i^{kT}(x_i - \overline{x}^k) + (\rho/2)\|x_i - \overline{x}^k\|_2^2 \right) \\ y_i^{k+1} &:= y_i^k + \rho(x_i^{k+1} - \overline{x}^{k+1}) \end{aligned}$$

where y is the dual variable, $\overline{x}^k$ is the average of the minimizers at step k, and ρ is the regularization constant

- Can converge in approximately 20 iterations

Inverse Sparse Covariance Selection

- The covariance of measurement noise can be estimated empirically using

$$\mathbb{E}(vv^T) = \frac{1}{N}\sum_{i}^{N}(y_i - h(x_i, u_i))(y_i - h(x_i, u_i))^T$$

- This matrix will almost certainly be dense, so if the true covariance is diagonal or sparse, then its inverse can be approximated by Inverse Sparse Covariance Selection formulated as

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(SX) - \log\det X + \lambda\|Z\|_1 \\ \text{subject to} & X - Z = 0 \end{array}$$

- Implemented in ADMM

$$\begin{aligned} X^{k+1} &:= \underset{X}{\operatorname{argmin}} \left(\mathbf{Tr}(SX) - \log\det X + (\rho/2)\|X - Z^k + U^k\|_F^2\right) \\ Z^{k+1} &:= S_{\lambda/\rho}(X^{k+1} + U^k) \\ U^{k+1} &:= U^k + (X^{k+1} - Z^{k+1}) \end{aligned}$$

Distributed Filtering

1. Compute empirical covariance matrix R
2. Compute sparse approximation of covariance inverse R^{-1} using ADMM
3. Decompose objective function into parallelizable components
4. Compute maximum likelihood estimation in parallel using ADMM