# STATE SPACE MODELS

JUST LIKE HMM BUT HIDDEN STATES ARE CONTINUOUS.

$$\begin{cases} z_T = g(u_T, z_{T-1}, \epsilon_T) \\ y_T = h(z_T, u_T, \delta_T) \end{cases}$$

- $z_T$ HIDDEN STATE
- $y_T$ OUTPUT/OBSERVATION
- $u_T$ INPUT/CONTROL
- $g$ TRANSITION MODEL
- $h$ OBSERVATION MODEL
- $\epsilon_T, \delta_T$ SYSTEM AND OBSERVATION NOISES

## TASKS

- ESTIMATE BELIEF STATE $P(z_T \mid y_{1-T}, u_{1-T}, \theta)$
- PREDICT FUTURE OBSERVABLES $P(y_{T+1} \mid y_{1-T})$

## LINEAR GAUSSIAN SSM / LG-SSM (LINEAR DYNAMICAL SYSTEM)

ALL CPD ARE LINEAR GAUSSIAN.

$$\begin{cases} z_T = A_T z_{T-1} + B_T u_T + \epsilon_T \\ y_T = C_T z_T + D_T u_T + \delta_T \\ \epsilon_T \sim N(0, Q_T) \\ \delta_T \sim N(0, R_T) \end{cases}$$

- IF $\theta = (A, B, C, D, Q, R)$ INDEPENDENT OF TIME $\longrightarrow$ STATIONARY
- COOL BECAUSE IF INITIAL BELIEF STATE IS GAUSSIAN $P(z_1) \sim N(\mu_{1|0}; \Sigma_{1|0})$ ALL FOLLOWING $z$ ALSO GAUSSIAN $P(z_T) \sim N(\mu_{t|t}; \Sigma_{t|t})$

## APPLICATIONS

- OBJECT TRACKING $z_T = (x_T, y_T, \dot{x}_T, \dot{y}_T)$ STATE VECTOR WITH POS/VEL PER COMPONENTS. $y_T = C_T z_T + \delta_T$. $C$ ZEROES OUT VELOCITIES. SEQUENTIAL BAYES UPDATES W/ KALMAN FILTER. CAN $P(z_{1T}, z_{2T} \mid y_{1T})$ WITH MARGINALIZATION. ESTIMATE LOCATION WITH POSTERIOR MEAN $E[z_T \mid y_{1T}]$
- ROBOTIC SLAM

  SLAM: SIMULTANEOUS LOCALIZATION AND MAPPING. $z_T = (x_T, L^{1-K})$ $x_T$ CUR POS, $L$ LANDMARKS. $x_T$ UNKNOWN. $y_T$ DISTANCE $x_T$ TO CLOSEST SET OF LANDMARKS. IF OBS AND MOTION MODELS ARE GAUSSIAN $\longrightarrow$ CAN KALMAN TO MAINTAIN BELIEF STATE. UNCERTAINTY GROWS OVER TIME BUT SHRINKS BACK WHEN AT FAMILIAR LOCATION $\longrightarrow$ CLOSING THE LOOP. POSTERIOR PRECISION MATRIX $\Lambda$ IS SPARSE. STARTS DIAGONAL BECAUSE ALL LANDMARKS START UNCORRELATED BUT BECOME CORRELATED BECAUSE ROBOT POSITION $\longleftrightarrow$ LANDMARK LOCATION. BECOME INTERDEPENDENT OVER TIME.

  - DYNAMICALLY PRUNE OUT WEAK EDGES
  - CONDITIONAL ON ROBOT PATH $x_{1:T}$, LANDMARK POSITIONS ARE INDEPENDENT. FASTSLAM.
- ONLINE INFERENCE FOR STATISTICAL MODELS

  HIDDEN STATE IS REGRESSION PARAMETERS. OBSERVATION MODEL IS CUR DATA VECTOR. UPDATE BELIEFS AS DATA COMES IN VIA KALMAN FILTER.

  RECURSIVE LEAST SQUARES $\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\sigma^2} \Sigma_{t|t} (y_t - x_t^T \hat{\theta}_{t-1}) x_t$ OPTIMAL STEP SIZE ADAPTATION AND CONVERGENCE IN ONE PASS
- TIME SERIES FORECASTING

  IDEA: CREATE GENERATIVE MODEL OF TS WITH LATENT PROCESSES. INTEGRATE OUT HIDDEN VARS TO OBTAIN PREDICTIONS OF VISIBLES

  GAUSSIAN-LINEAR MODEL $\longrightarrow$ ADDITIVE EFFECTS

  - LOCAL LEVEL MODEL: $\begin{cases} y_t = a_t + \epsilon_t^y & \epsilon_t^y \sim N(0, R) \\ a_t = a_{t-1} + \epsilon_t^a & \epsilon_t^a \sim N(0, Q) \end{cases}$
  - LOCAL LINEAR TREND: $\begin{cases} y_t = a_t + \epsilon_y \\ a_t = a_{t-1} + b_{t-1} + \epsilon_a \\ b_t = b_{t-1} + \epsilon_b \end{cases}$ UNROLLING BECOMES $a_t = a_0 + b_0 t$
  - ARMA MODELS CAN BE SEEN AS SSM / MARKOV CHAINS
  - SEASONALITY: ADD LATENT PROCESS OF OFFSET TERMS SUMMING TO 0 ON AVG OVER $|S|$ STEPS CYCLE

# INFERENCE FOR LG-SSM: KALMAN FILTER

ONLINE → ANALOGOUS TO FWD ALGO FOR HMM

OFFLINE → ANALOGOUS TO FWD-BWD FOR HMM

## KALMAN FILTER

EXACT BAYESIAN FILTERING FOR LG-SSM. MARGINAL POSTERIOR $p(z_t|y_{1:t}, u_{1:t}) = N(z_t|\mu_t, \Sigma_t)$

- PREDICTION STEP

$$\begin{cases} p(z_t|y_{1:T}, u_{1:T}) = \int N(z_t|A_T z_{T-1} + B_T u_T, Q_T) N(z_{t-1}|\mu_{T-1}, \Sigma_{T-1}) dz_{T-1} = N(z_t|\mu_{T|T-1}, \Sigma_{T|T-1}) \\ \mu_{T|T-1} = A_t \mu_{t-1} + B_t u_t \\ \Sigma_{T|T-1} = A_T \Sigma_{T-1} A_t^T + Q_t \end{cases}$$

- MEASUREMENT STEP

$$\begin{cases} p(z_t|y_{1:t}, u_t) = N(z_T|\mu_t, \Sigma_t) \\ \mu_t = \mu_{t|t-1} + K_t R_t \\ \Sigma_t = (I - K_t C_t)\Sigma_{t|t-1} \end{cases}$$

- $R_t = y_t - \hat{y}_t$
- $\hat{y}_t = C_t \mu_{t|t-1} + D_t u_t$
- $K_t = \Sigma_{t|t-1} C_t^T S_t^{-1}$ KALMAN GAIN

- $S_t = cov[R_t|y_{1:T-1}, u_{1:T}] = C_t \Sigma_{t|t-1} C_t^T + R_t$
- IF STRONG PRIOR OR NOISY SENSORS $|K_t|$ SMALL, LITTLE WEIGHT ON CORRECTION TERM

- MARGINAL LIKELIHOOD

$$\log p(y_{1:T}|u_{1:T}) = \sum \log p(y_t|y_{1,T-1}, u_{1,t}) \longrightarrow N(y_t|C_t \mu_{t|t-1}, S_t)$$

- POSTERIOR PREDICTIVE

$$p(y_t|y_{1,t-1}, u_{1,t}) = N(y_t|C\mu_{t|t-1}, C\Sigma_{t|t-1} C^T + R)$$

TIMESERIES FORECASTING

- COMPUTATIONAL ISSUES

MATRIX INVERSION FOR $K_t = O(|y_t|^3)$. MATRIX MULT FOR $\Sigma_t$ IS $(|z_t|^2)$. $K_t$ CAN BE PRECOMPUTED BECAUSE NO DEPEND ON OBSERVATIONS BECAUSE LGS. $\Sigma_t$ CAN BE ITERATIVELY UPDATED WITH RICCATI EQUATIONS. CONVERGES FOR STATIONARY SYSTEMS.

IN PRACTICE: FOR NUMERICAL STABILITY INFORMATION FILTER FORM, UPDATES CANONICAL PARAMETERS $\Lambda_t = \Sigma_t^{-1}$, $\eta_t = \Lambda_t \mu_t$ ONLINE VS MOMENT PARAMETERS. SQUARE ROOT FILTER VIA CHOLESKY DECOMPOSITION

## KALMAN SMOOTHING

OFFLINE. CONDITION ON PAST + FUTURE DATA. $p(z_t|y_{1,T})$. REDUCES UNCERTAINTY. RTS/KALMAN SMOOTHER

- FIRST DO KALMAN FILTER → 'MESSAGE PASSING ON GRAPH' L → R. OBTAIN $p(z_T|y_{1,T})$
- THEN WORK BACKWARDS AND COMBINE

$$\begin{cases} p(z_t|y_{1,T}) = N(\mu_{t|T}, \Sigma_{t|T}) \\ \mu_{t|T} = \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t}) \\ \Sigma_{t|T} = \Sigma_{t|t} + J_t(\Sigma_{t+1|T} - \Sigma_{t+1|t}) J_t^T \end{cases}$$

- $J_t = \Sigma_{t|t} A_{t+1}^T \Sigma_{t+1|t}^{-1}$ BACKWARD KALMAN GAIN

- INITIALIZES FROM $\mu_{T|T}$, $\Sigma_{T|T}$ FROM KF
- DOES NOT ACCESS DATA $y_{1:T}$, CAN THROW AWAY.

- UNLIKE FWD/BWD KALMAN NEEDS FWD PASS TO COMPUTE BWD PASS; CAN REFORMULATE BUT SUPER SBATTI AND INCONVENIENT BECAUSE NEED DATA, BACKWARD MSG IS LIKELIHOOD, NOT POSTERIOR; EXACT INFERENCE NOT POSSIBLE, LIKELIHOOD NO BELLO.

# LEARNING FOR LG-SSM

AKA SYSTEMS IDENTIFICATION FROM THE '60s. IN IE TIME SERIES C AND A ARE KNOWN AND FIXED. ONLY Q, R ARE TO BE LEARNED. CAN OFFLINE OR XACT POSTERIOR $p(z_t, R, Q | y_{1,t})$ - IS NORMAL INVERSE WISHART

- WHEN WE HAVE TO LEARN A AND C TOO SET $Q = I$, R DIAG WLOG. IMPOSE EIGENVALUES OF A $\lambda < 1$ TO AVOID BLOWUPS.

## - FULL OBSERVABLE DATA

MLE OR FULL POSTERIOR. MULTIVARIATE LINEAR REGRESSION PROBLEM $z_{t-1} \to z_t$ ; $z_t \to y_t$ $J(A) = \sum_t (z_t - A z_{t-1})^2$. SAME FOR C.

Q FROM RESIDUALS OF $z_t$ FROM $z_{t-1}$. R FROM RESIDUALS $y_t$ FROM $z_t$.

## - NON FULL OBS.

ONLY OUTPUT SEQUENCE. • EM. BAUM-WELCH WITH KALMAN SMOOTHING AT BACKWARD PASS.

- EM A VOLTE FA MERDA, SUBSPACE METHOD: ASSUME LITTLE NOISE, $y_t = CA_{t-1} z_1$. ALL OBSERVATIONS FROM $DIM(z_t)$ MANIFOLD/SUBSPACE CAN IDENTIFY WITH PCA; THEN USE $z_t$s TO FIT MODEL OR INITIALIZE EM.
- BAYESIAN METHODS: VAR BAYES, GIBBS SAMPLING.

# APPROX-ONLINE INFERENCE FOR NONLINEAR/NONGAUSSIAN SSM

IE STUFF DOES NOT MOVE IN STRAIGHT LINE, OR PARAMS $\theta$ UNKNOWN AND NONLINEAR HAPPENS WHEN ADDED TO STATESPACE. OR NONGAUSSIAN NOISE.

$\longrightarrow$ GENERALLY APPROXIMATE POSTERIOR BY A GAUSSIAN, USE 1ST ORDER APPROX BY A GAUSSIAN, USE EXACT $f$ BUT PROJECT $f(x)$ ONTO SPACE OF GAUSSIANS

## EXTENDED KALMAN FILTER

NON LINEAR MODEL BUT NOISE STILL GAUSSIAN. LINEARIZE $g, h$ WITH 1ST ORDER TAYLOR XPANSION. THEN USE STD KALMAN FILTER.
WE APPROX STATIONARY NONLINEAR WITH NONSTATIONARY LINEAR. GAUSSIAN $\longrightarrow$ NONLINEARITY, MC APPROX OF $E[g]$, $VAR[g]$; EKF EVALUATES LINEARIZED $g$ AT CURRENT MODE $\mu \longrightarrow$ PASS GAUSSIAN THROUGH THIS APPROX. $p(y_t|z_t) \approx N(y_t | h(\mu_{t|t-1}) + H_t(y_t - \mu_{t|t-1}), R_t)$

$H_t$: JACOBIAN AT PRIOR MODE

CAN IMPROVE BY RELINEARIZATIONS OF EQUATIONS $\longleftarrow$ ITERATED EKF

- WORKS POORLY WHEN PRIOR COVARIANCE LARGE, MASS IS SPREAD TOO MUCH WHERE IT'S NOT RELEVANT. OR WHEN HIGHLY NONLINEAR NEAR CURRENT MEAN

## UNSCENTED KALMAN FILTER

IS BETTER EKF. PASS A DETERMINISTICALLY CHOSEN SET OF POINTS THROUGH NONLINEARITY (SIGMA POINTS) AND FIT A GAUSSIAN TO TRANSF. POINTS.
DOES NOT REQUIRE COMPUTING DERIVATIVES OR JACOBIANS (YAY!). ACCURATE TO AT LEAST II ORDER. $O(d^3)$ OPERATIONS PER TIMESTEP. $d$ IS LATENT SPACE SIZE.

- UNSCENTED TRANSFORM

$2d+1$ SIGMA POINTS. $x = \{\mu, \{\mu + (\sqrt{(d+\lambda)\Sigma})\}_{i=1}^{d}, \{\mu - (\sqrt{(d+\lambda)\Sigma})\}_{i=1}^{d}\}$. $\lambda = \alpha^2(d+k) - d$. SCALING PARAM.

SIGMA GO THROUGH NONLINEARITY. ESTIMATE $\mu_y, \Sigma_y$ WITH WEIGHTS DEPENDING ON $\lambda$. $\alpha, \beta, k$ PROBLEM DEPENDANT. IN 1d $\lambda = 2$, $\sigma: \{\mu, \mu \pm \sigma\sqrt{3}\}$

- ALGO

TWO APPLICATIONS OF TRANSFORM. $p(z_t | y_{1,t-1}, u_{1,t})$, $p(z_t | y_{1,t}, u_{1,t})$

## ASSUMED DENSITY FILTERING:

EXACT UPDATE STEP BUT USE CONVENIENT APPROXIMATION OF POSTERIOR. UNKNOWNS ARE $\theta$. Q IS SET OF TRACTABLE DISTRIBUTIONS. $q_{t-1}(\theta_{t-1}) \approx p(\theta_{t-1} | y_{1,t-1})$

$\hat{p}(\theta_t) = \frac{1}{Z_t} p(y_t|\theta_t) q_{t|t-1}(\theta_t)$, $Z_t$ NORMALIZATION CONSTANT $= \int p(y_t|\theta_t) q_{t|t-1}(\theta_t) d\theta_t$ ; $q_{t|t-1}(\theta_t) = \int p(\theta_t|\theta_{t-1}) q_{t-1}(\theta_{t-1}) d\theta_{t-1}$ 1 STEP AHEAD PREDICTIVE

- IF $\hat{p}(\theta_t) \notin Q$ SEEK BEST TRACTABLE APPROX $q(\theta_t) = \text{ARGMIN } KL(\hat{p}(\theta_t) || q(\theta_t))$. PROJECTION ONTO SPACE OF TRACTABLE DISTRIBUTIONS. PREDICT - UPDATE - PROJECT CYCLE
- IF Q IN EXPONENTIAL FAMILY $\to$ KL MINIMIZATION VIA MOMENT MATCHING.
- EXAMPLES: BOYEN-KOLLER ALGO FOR DBN, TRUESKILL FOR XBOX RANKING/MATCHMAKING (ONLINE INFERENCE FOR MASSIVE SHIT)

# HYBRID SSMS

CONTAIN BOTH DISCRETE AND CONTINUOUS HIDDEN VARIABLES. HMM + LG-SSM FOR INSTANCE. = LINEAR SWITCHING DYNAMICAL SYSTEM (SLDS), JUMP MARKOV LINEAR SYSTEM (JMLS), SWITCHING STATE SPACE MODEL (SSSM)

$q_t$ DISCRETE LV; $z_t$ CONTINUOUS LV; $y_t$ OBSERVATION, OPTIONAL $u_t$ CONTROL.

$$\begin{cases} p(q_t = k \mid q_{t-1} = j, \theta) = A_{ij} \\ p(z_t \mid z_{t-1}, q_t = k, u_t, \theta) = \mathcal{N}(z_t \mid A_k z_{t-1} + B_k u_t, Q_k) \\ p(y_t \mid z_t, q_t = k, u_t, \theta) = \mathcal{N}(y_t \mid C_k z_{t-m} + D_k u_t, R_k) \end{cases}$$

- INFERENCE IS INTRACTABLE → EXPONENTIAL EXPLOSION IN NUMBER OF MODES PER FUTURE TIMESTEPS DUE TO DISCRETE VAR.
  - APPROX USING MULTIPLE HYPOTHESIS TRACKING (PRUNING), MONTECARLO (SAMPLE + FILTER), USE ADF

## ADF APPROXIMATION

APPROXIMATES EXPONENTIALLY LARGE GAUSSIAN MIXTURE WITH SMALLER ONE. **GAUSSIAN SUM FILTER**. RUNS K KALMANS IN PARALLEL, EACH ONE IS MIXTURE. GENERALIZED PSEUDO BAYES FILTER 2ND ORDER **GPB2**. $b_{t-1}$ IS MIXTURE OF $K$ GAUSSIANS. ONE x DISCRETE STATE. FIT MODELS. COLLAPSE GAUSSIANS TO SINGLE MIXTURE. MINIMIZING KL. WEAK MARGINALIZATION. $K^2$ FILTERS AT EACH STEP.
MIXTURES

- **ELSE** REPRESENT BELIEF STATE BY SINGLE GAUSSIAN AND MARGINALIZE DISCRETE SWITCH AT EACH STEP.

## APPLICATIONS.

**DATA ASSOCIATIONS / MULTI OBJECT TRACKING** K OBJECTS SIMULTANEOUSLY. CAN HAVE MISSED DETECTIONS + FALSE ALARMS. $y_{tk}$ OBSERVATIONS; $z_{tk}$ MAPPINGS BETWEEN OBJECTS AND DETECTIONS. SQUARE WEIGHT COMPOSITE MATRIX. USE A NEAREST NEIGHBOR (IN TIME) DATA ASSOCIATION HEURISTIC. WEIGHTED BIPARTITE MATCHING. KALMAN UPDATE.

**FAULT DIAGNOSIS** FOR INDUSTRIAL PLANTS.

**ECONOMETRICS FORECASTING** FOR REGIMES / TRENDS