Uncertainty propagation

- Moment matching for <u>deterministic propagation</u> of the mean $\mu(s_t)$ and the covariance $\Sigma(s_t)$ of the distribution of dynamics-reward model
- The immediate performance measure is:

$$\mathbb{E}[r(s_t, a_t)] = \int r(s_t, a_t)\mathcal{N}(s_t \mid \mu_t, \Sigma_t)ds_t$$

- If reward not emitted - formulated as polynomial function

Fast optimisation

- From PMP a sequence of a constraint optimisation for each time step
- Dynamics-Lagrangian-multipliers in closed-form, Hamiltonian gradient same as Reward gradient
- Optimisation (analytical) up to (second) order in dynamics-reward model
- State and action constraints (analytical) up to second order
- “An interior point algorithm for large-scale nonlinear programming” - “trust-constr” used for experiments (we use BFGS)