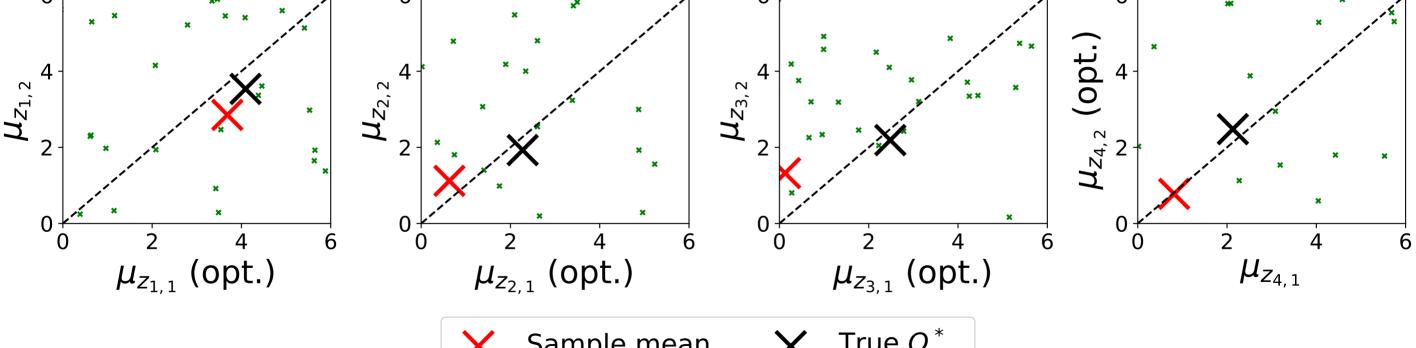UBE samples PriorMDP ($t = 500$, $\zeta = 1.0$, $N_s = 4$, $N_a = 2$)

Sample mean     True $Q^*_{\mathbf{s}, \mathbf{a}}$