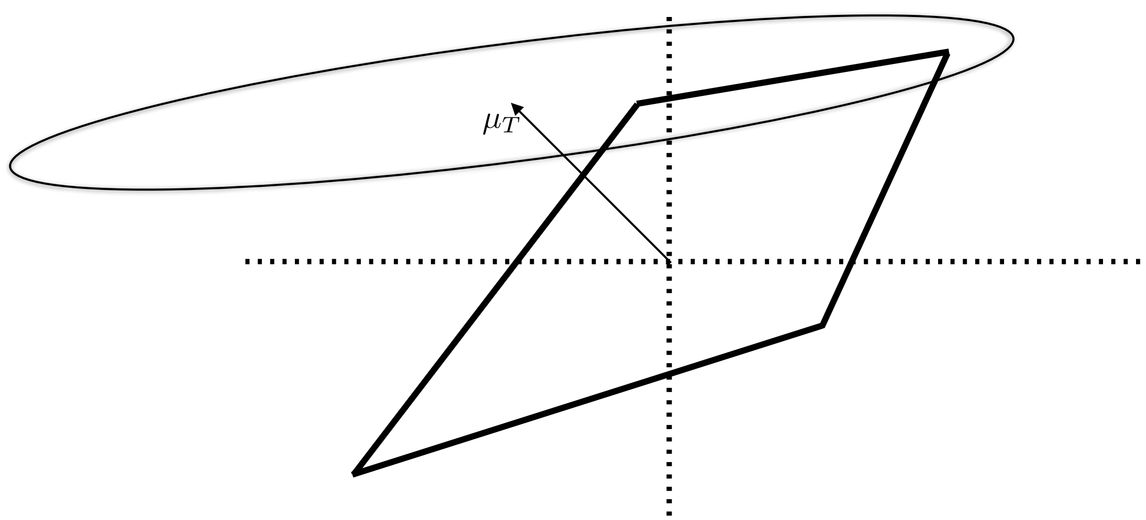
Online LP with Gaussian Prior

- Linear program $\max_{Ax \leq b} \theta^\top x$
- Gaussian prior $\theta \sim \mathcal{N}(\mu_0, \Sigma_0)$
- Observation $R_t = \theta^\top X_t + W_t \quad W_t \sim \mathcal{N}(0, 1)$
- Bayesian update ≅ linear regression

$$\mu_T = \arg\min_{\hat{\theta}} \sum_{t=1}^{T} (R_t - \hat{\theta}^\top X_t)^2 + (\hat{\theta} - \mu_0)^\top \Sigma_0^{-1} (\hat{\theta} - \mu_0)$$

$$(\theta - \mu_T | \mathbb{F}_T) \sim \mathcal{N}(0, \Sigma_T)$$



Bayes Optimal Solution

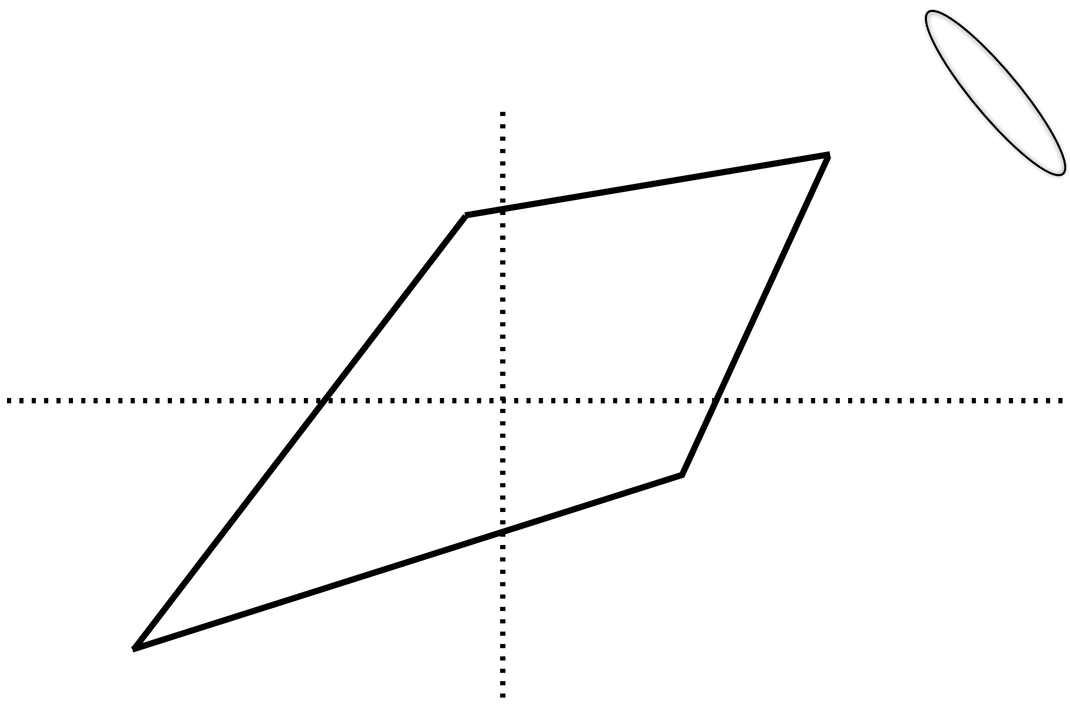
- Finite horizon objective

$$\max_{\pi_1,\ldots,\pi_T} \mathbb{E}\left[\sum_{t=1}^{T} \phi^\top \pi_t(\mu_t, \Sigma_t)\right]$$

- Dynamic programming
  - State : (μ_t, Σ_t)
  - Action : $X_t = \pi_t(\mu_t, \Sigma_t)$
  - Intractable
- Resort to heuristics

ε-greedy Exploration Schemes

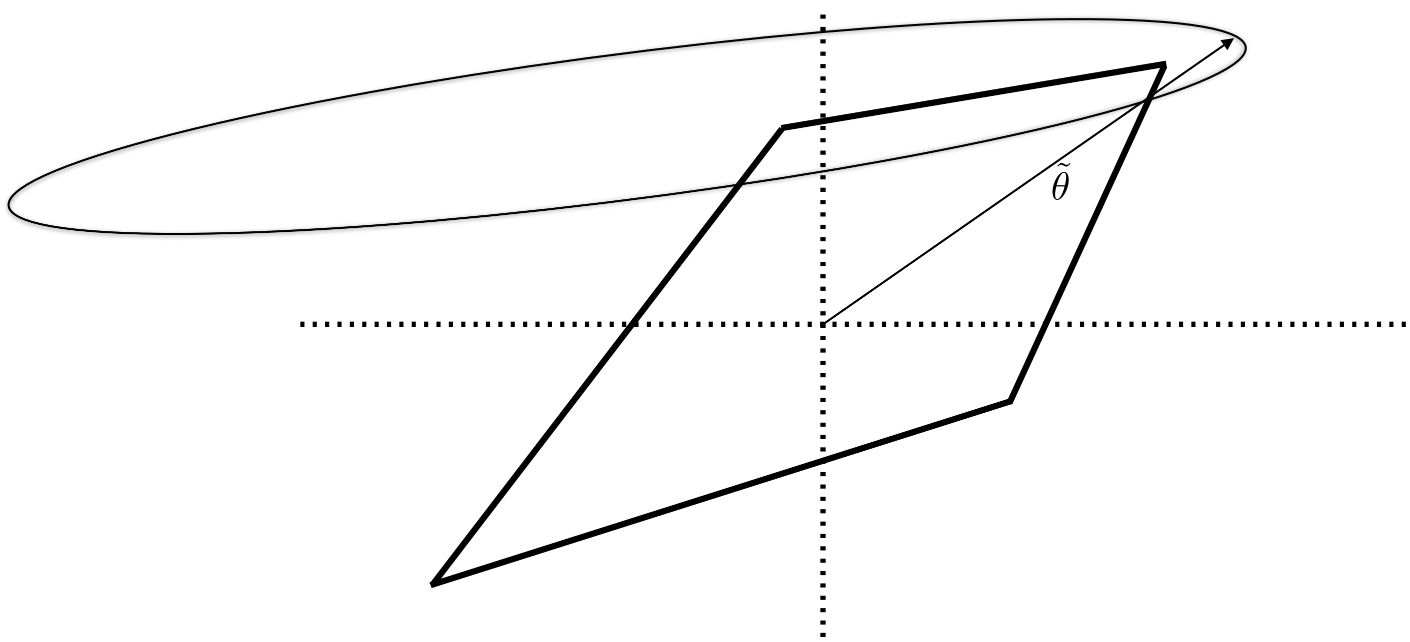
- uniform sampling
  - exploit: maximize expected reward
  - explore: choose randomly from possible optima



- informative sampling
  - exploit: maximize expected reward
  - explore: choose most informative action

Upper-Confidence Bounds

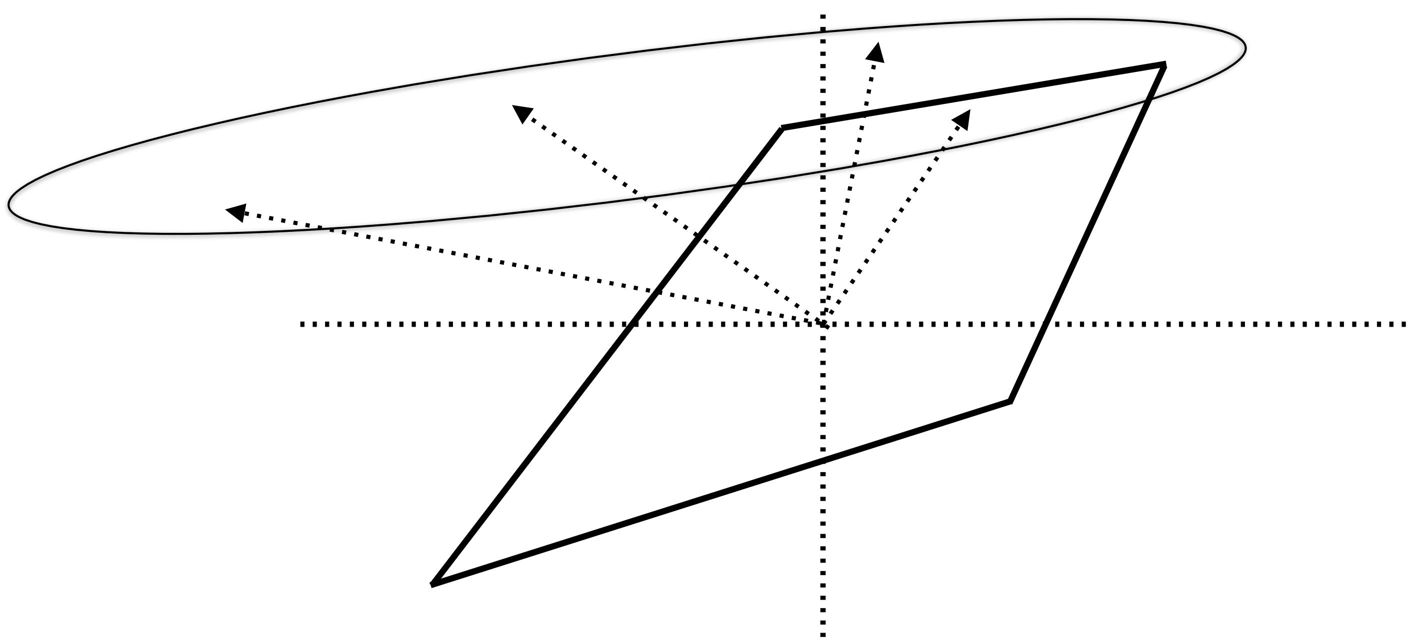
- Maintain confidence set Θ_t
  - Set of statistically plausible models
- Optimistic optimization $\max_{Ax \leq b} \max_{\tilde{\theta} \in \Theta_t} \tilde{\theta}^\top x$



- Chooses only plausibly optimal actions
  - Avoids exploring when not helpful
- Either
  - Exploit: near-optimal performance
  - Explore: reduce uncertainty about plausible actions

Thompson Sampling

- Sample model from posterior $\tilde{\theta} \sim p_{t-1}$
- Optimize for that sample $\max_{Ax \leq b} \tilde{\theta}^\top x$



- Optimistic?
- "Randomized approximation" of UCB

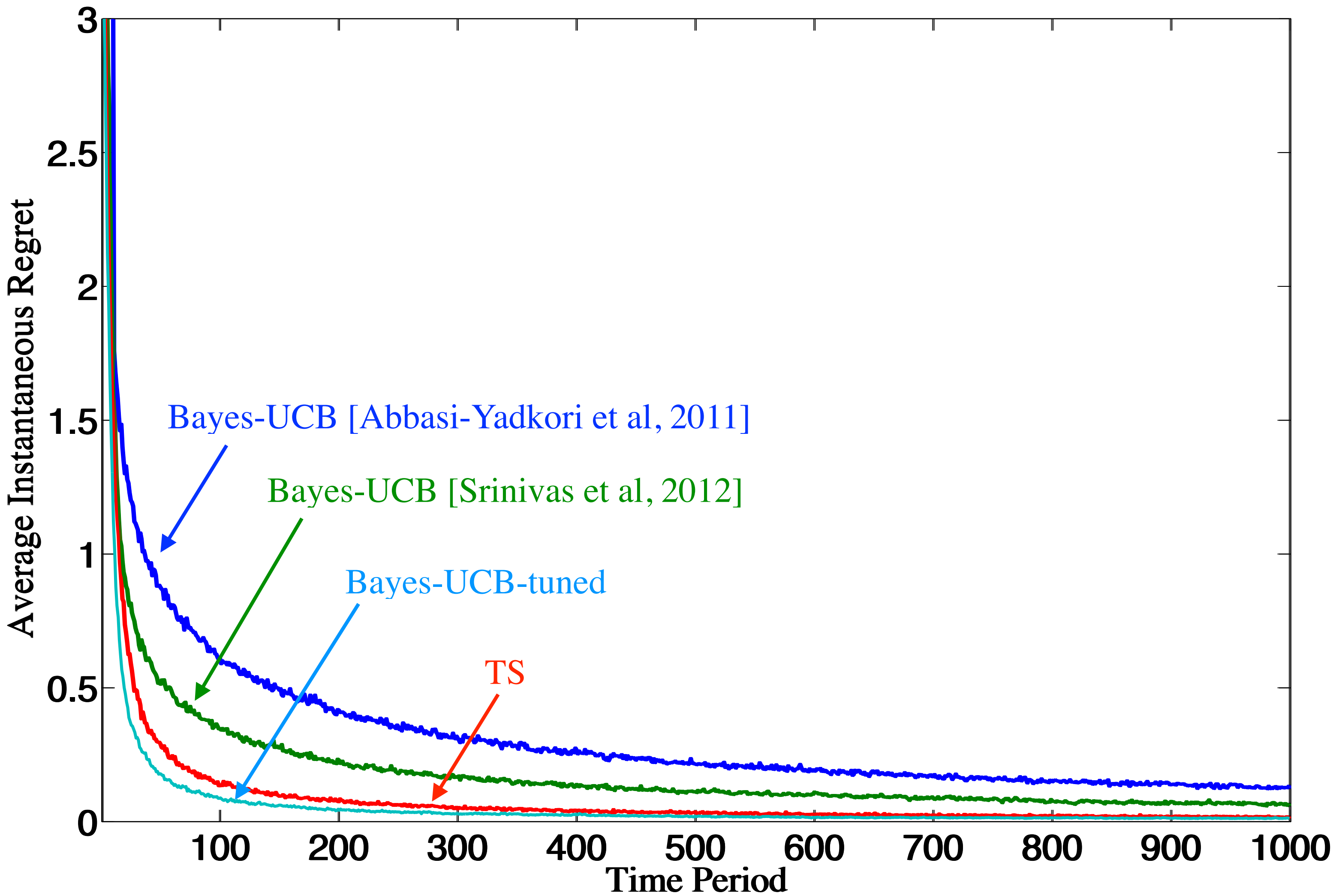
Regret

- Instantaneous regret $f_\theta(x^*) - R_t$
- Expected regret

$$\mathbb{E}\left[f_\theta(x^*) - R_t\right] = \mathbb{E}\left[f_\theta(x^*) - f_\theta(X_t)\right]$$

- Expected cumulative regret

$$\sum_{t=1}^{T} \mathbb{E}\left[f_\theta(x^*) - f_\theta(X_t)\right]$$

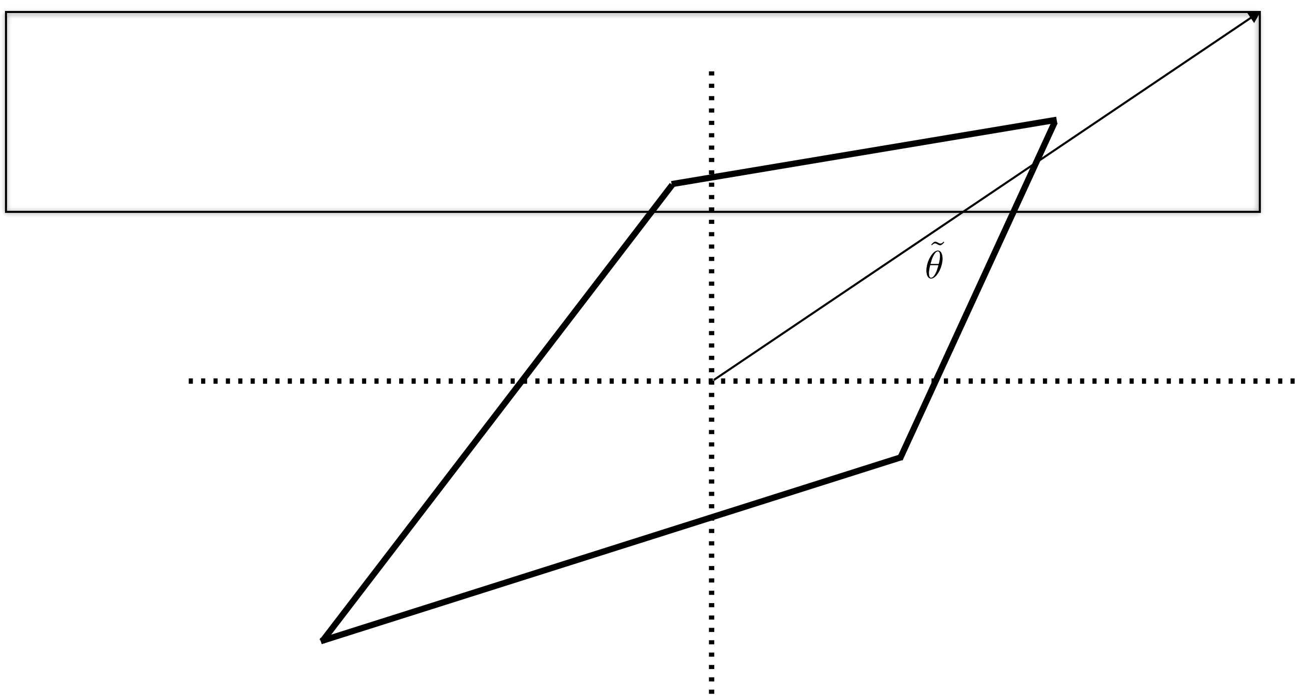


Computational Considerations

- For LP-Gaussian problem
  - Ellipsoidal confidence set makes optimization intractable

$$\max_{Ax \leq b} \max_{\tilde{\theta} \in \Theta_t} \tilde{\theta}^\top x$$

  - Use hyper-rectangular confidence sets?
    - Regret increases by factor of d
    - Implicit independence too conservative
  - TS expected regret ≅ "Bayes-UCB"



- More broadly
  - Bayes-UCB sometimes intractable but typically not
  - TS provides a computationally efficient approximation

General Online Optimization

- General information structures

$$X_t \longrightarrow \boxed{\text{system}} \longrightarrow Y_t \qquad R_t = R(X_t)$$

$$f_\theta(X_t) = \mathbb{E}[R_t | \theta, X_t]$$

- UCB

$$\max_{x \in \mathcal{X}} \max_{\tilde{\theta} \in \Theta_t} f_{\tilde{\theta}}(x)$$

- Thompson Sampling

$$\tilde{\theta} \sim p_t \qquad \max_{x \in \mathcal{X}} f_{\tilde{\theta}}(x)$$

- Time-dependent action constraints

$$\max_{x \in \mathcal{X}_t} \max_{\tilde{\theta} \in \Theta_t} f_{\tilde{\theta}}(x) \qquad \max_{x \in \mathcal{X}_t} f_{\tilde{\theta}}(x)$$

  - Context
  - Adversaries
  - Caution