An Introduction to Bayesian Optimisation and (Potential) Applications in Materials Science

Kirthevasan Kandasamy
Machine Learning Dept, CMU

Electrochemical Energy Symposium
Pittsburgh, PA, November 2017
Designing Electrolytes in Batteries

Electrolyte Experiment

Solvent       Salt
$x_1$: %EC    $x_4$: LiPF$_6$
$x_2$: %DMC    $x_5$: LiNO$_3$
$x_3$: %EMC

$\mathbf{x} \rightarrow f(\mathbf{x})$

Conductivity of electrolyte
Black-box Optimisation in Computational Astrophysics

$x$ → Cosmological Simulator → Observation → Likelihood computation → $f(x)$

E.g:
- Hubble Constant
- Baryonic Density

Likelihood Score
Black-box Optimisation

Expensive Blackbox Function

Other Examples:
- Pre-clinical Drug Discovery
- Optimal policy in Autonomous Driving
- Synthetic gene design
Black-box Optimisation

\( f : \mathcal{X} \rightarrow \mathbb{R} \) is an expensive, black-box function, accessible only via noisy evaluations.
Black-box Optimisation

\( f : \mathcal{X} \rightarrow \mathbb{R} \) is an expensive, black-box function, accessible only via noisy evaluations.
Black-box Optimisation

\( f : \mathcal{X} \to \mathbb{R} \) is an expensive, black-box function, accessible only via noisy evaluations.
Let \( x_\star = \arg\max_x f(x) \).
Outline

- Part I: Bayesian Optimisation
  - Bayesian Models for $f$
  - Two algorithms: upper confidence bounds & Thompson sampling

- Part II: Some Modern Challenges
  - Multi-fidelity Optimisation
  - Parallelisation
Bayesian Models for $f$ e.g. Gaussian Processes ($GP$)

$GP$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$. 

Bayesian Models for $f$ e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Functions with no observations

$$f(x)$$

After $t$ observations, $f(x) \sim \mathcal{N}(\mu_t(x), \sigma^2_t(x))$. 
Bayesian Models for $f$ e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Prior $\mathcal{GP}$

After $t$ observations, $f(x) \sim N(\mu_t(x), \sigma^2_t(x))$. 

$\mu_t(x)$ and $\sigma^2_t(x)$ are the mean and variance functions, respectively.
Bayesian Models for $f \quad \text{e.g. Gaussian Processes (GP)}$

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Observations

$$f(x)$$

After $t$ observations, $f(x) \sim \mathcal{N}(\mu_t(x), \sigma^2_t(x))$. 
Bayesian Models for $f$ e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Posterior $\mathcal{GP}$ given observations
Bayesian Models for $f$  

**$\mathcal{GP}$**: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Posterior $\mathcal{GP}$ given observations

After $t$ observations,  

$$f(x) \sim \mathcal{N}(\mu_t(x), \sigma_t^2(x)).$$
Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)  
(Srinivas et al. 2010)

\begin{align*}
1) & \text{Construct posterior } \mathcal{GP} . \\
2) & \phi_t = \mu_t - 1 + \beta \frac{1}{t} \sigma_t - 1 \text{ is a UCB.} \\
3) & \text{Choose } x_t = \arg\max_x \phi_t (x) . \\
4) & \text{Evaluate } f \text{ at } x_t.
\end{align*}
Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)

1) Construct posterior $\mathcal{GP}$.

$\phi_t = \mu_t - \frac{1}{t} + \frac{\beta}{2t} \sigma_t - \frac{1}{t}$ is a UCB.

Choose $x_t = \arg\max_x \phi_t(x)$.

Evaluate $f$ at $x_t$. 
Bayesian Optimisation with Upper Confidence Bounds

Model \( f \sim \mathcal{GP} \).

Gaussian Process Upper Confidence Bound (GP-UCB)  
(Srinivas et al. 2010)

1) Construct posterior \( \mathcal{GP} \).  
2) \( \varphi_t = \mu_{t-1} + \beta_t^{1/2} \sigma_{t-1} \) is a UCB.
Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB) 

(Srinivas et al. 2010)

1) Construct posterior $\mathcal{GP}$.  
2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2} \sigma_{t-1}$ is a UCB.  
3) Choose $x_t = \text{argmax}_x \varphi_t(x)$. 

$f(x)$

\begin{align*}
\varphi_t &= \mu_{t-1} + \beta_t^{1/2} \sigma_{t-1} \\
\end{align*}
Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB) 

(Srinivas et al. 2010)

1) Construct posterior $\mathcal{GP}$.  
2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2} \sigma_{t-1}$ is a UCB.  
3) Choose $x_t = \arg\max_x \varphi_t(x)$.  
4) Evaluate $f$ at $x_t$.  

$f(x)$

$\varphi_t = \mu_{t-1} + \beta_t^{1/2} \sigma_{t-1}$

$\varphi_t$
GP-UCB  (Srinivas et al. 2010)

\[ f(x) \]

\[ t = 1 \]
GP-UCB  (Srinivas et al. 2010)

\[ f(x) \]

\[ t = 2 \]

\[ x \]
$f(x)$

$t = 3$

$x$
GP-UCB \hspace{1cm} (Srinivas et al. 2010)

\[ t = 4 \]

\[ x \]

\[ f(x) \]
GP-UCB (Srinivas et al. 2010)

\[ f(x) \]

\[ t = 5 \]
GP-UCB (Srinivas et al. 2010)

t = 6

\[ f(x) \]

\[ x \]

\[ t = 6 \]
GP-UCB (Srinivas et al. 2010)

\[ t = 7 \]

\[ f(x) \]

\[ x \]

\( t = 7 \)
GP-UCB  (Srinivas et al. 2010)
GP-UCB (Srinivas et al. 2010)

\[ f(x) \]

\[ t = 25 \]

\[ x \]
Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(0, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

1. Construct posterior $\mathcal{GP}$.
2. Draw sample $g$ from posterior.
3. Choose $x_t = \text{argmax}_x g(x)$.
4. Evaluate $f$ at $x_t$.
Bayesian Optimisation with Thompson Sampling

Model $f \sim GP(0, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

1) Construct posterior $\mathcal{GP}$.

\begin{equation*}
\text{Thompson Sampling (TS)}
\end{equation*}
Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(0, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

1) Construct posterior $\mathcal{GP}$.  
2) Draw sample $g$ from posterior.
Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(0, \kappa)$.

Thompson Sampling (TS) 

1) Construct posterior $\mathcal{GP}$. 
2) Draw sample $g$ from posterior. 
3) Choose $x_t = \text{argmax}_x g(x)$.

(Thompson, 1933)
Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(0, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

1) Construct posterior $\mathcal{GP}$.  
2) Draw sample $g$ from posterior.  
3) Choose $x_t = \arg\max_x g(x)$.  
4) Evaluate $f$ at $x_t$. 

$\begin{align*}
\text{Model } f &\sim \mathcal{GP}(0, \kappa) . \\
\text{Thompson Sampling (TS)} &\quad \text{(Thompson, 1933).} \\
\text{1) Construct posterior } \mathcal{GP}. \\
\text{2) Draw sample } g \text{ from posterior.} \\
\text{3) Choose } x_t = \arg\max_x g(x). \\
\text{4) Evaluate } f \text{ at } x_t. 
\end{align*}$
More on Bayesian Optimisation

Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of $f$. 

- Expected improvement (Jones et al. 1998)
- Probability of improvement (Kushner et al. 1964)
- Predictive entropy search (Hernández-Lobato et al. 2014)
- Information directed sampling (Russo & Van Roy 2014)

Other Bayesian models for $f$:
- Neural networks (Snoek et al. 2015)
- Random Forests (Hutter 2009)
Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of $f$.

Other criteria for selecting $x_t$:

- Expected improvement (Jones et al. 1998)
- Probability of improvement (Kushner et al. 1964)
- Predictive entropy search (Hernández-Lobato et al. 2014)
- Information directed sampling (Russo & Van Roy 2014)
More on Bayesian Optimisation

Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of $f$.

Other criteria for selecting $x_t$:

- Expected improvement (Jones et al. 1998)
- Probability of improvement (Kushner et al. 1964)
- Predictive entropy search (Hernández-Lobato et al. 2014)
- Information directed sampling (Russo & Van Roy 2014)

Other Bayesian models for $f$:

- Neural networks (Snoek et al. 2015)
- Random Forests (Hutter 2009)
Some Modern Challenges/Opportunities

1. Multi-fidelity Optimisation (Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

2. Parallelisation (Kandasamy et al. Arxiv 2017)
1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

Desired function $f$ is very expensive, but ... we have access to cheap approximations.
1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

Desired function $f$ is very expensive, but ... we have access to cheap approximations.

$f_1, f_2, f_3 \approx f$ which are cheaper to evaluate.
1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

Desired function $f$ is very expensive, but we have access to cheap approximations.

$\mathbf{E.g.}$ $f$: a real world battery experiment
$f_2$: lab experiment
$f_1$: computer simulation

$f_1, f_2, f_3 \approx f$ which are cheaper to evaluate.
MF-GP-UCB

Multi-fidelity Gaussian Process Upper Confidence Bound

With 2 fidelities (1 Approximation),

Theorem: MF-GP-UCB finds the optimum $x^*$ with less resources than GP-UCB on $f(2)$.

Can be extended to multiple approximations and continuous approximations.
MF-GP-UCB

(Kandasamy et al. NIPS 2016b)

Multi-fidelity Gaussian Process Upper Confidence Bound

With 2 fidelities (1 Approximation),

**Theorem:** MF-GP-UCB finds the optimum $x_*$ with less resources than GP-UCB on $f^{(2)}$. 
**Theorem:** MF-GP-UCB finds the optimum \( x_* \) with less resources than GP-UCB on \( f^{(2)} \).

Can be extended to multiple approximations and continuous approximations.
**Experiment:** Cosmological Maximum Likelihood Inference

- Type Ia Supernovae Data

- Maximum likelihood inference for 3 cosmological parameters:
  - Hubble Constant $H_0$
  - Dark Energy Fraction $\Omega_\Lambda$
  - Dark Matter Fraction $\Omega_M$

- Likelihood: Robertson Walker metric (Robertson 1936)

Requires numerical integration for each point in the dataset.
Experiment: Cosmological Maximum Likelihood Inference

3 cosmological parameters.

Fidelities: integration on grids of size \((10^2, 10^4, 10^6)\).  

\(d = 3\)  
\(M = 3\)
Experiment: Hartmann-3D

2 Approximations (3 fidelities).
We want to optimise the $m = 3^{rd}$ fidelity, which is the most expensive. $m = 1^{st}$ fidelity is cheapest.

![Query frequencies for Hartmann-3D](image)
2. Parallelising function evaluations

Parallelisation with $M$ workers: can evaluate $f$ at $M$ different points at the same time.

E.g.: Test $M$ different battery solvents at the same time.
2. Parallelising function evaluations

Parallelisation with $M$ workers: can evaluate $f$ at $M$ different points at the same time.
E.g.: Test $M$ different battery solvents at the same time.

Sequential evaluations with one worker
2. Parallelising function evaluations

Parallelisation with $M$ workers: can evaluate $f$ at $M$ different points at the same time.
E.g.: Test $M$ different battery solvents at the same time.

Sequential evaluations with one worker

Parallel evaluations with $M$ workers (Asynchronous)
2. Parallelising function evaluations

Parallelisation with \( M \) workers: can evaluate \( f \) at \( M \) different points at the same time.

E.g.: Test \( M \) different battery solvents at the same time.

Sequential evaluations with one worker

Parallel evaluations with \( M \) workers (Asynchronous)

Parallel evaluations with \( M \) workers (Synchronous)
Parallel Thompson Sampling

(Kandasamy et al. Arxiv 2017)

**Asynchronous: asyTS**

At any given time,

1. \((x', y') \leftarrow\) Wait for a worker to finish.
2. Compute posterior \(GP\).
3. Draw a sample \(g \sim GP\).
4. Re-deploy worker at \(\text{argmax } g\).

\[1\ 4\ 9\ 12\ \ldots\]
\[2\ 6\ 8\ 10\ \ldots\]
\[3\ 5\ 7\ 11\ \ldots\]

Time →
Parallel Thompson Sampling
(Kandasamy et al. Arxiv 2017)

Asynchronous: \textit{asyTS}

At any given time,
1. \((x', y') \leftarrow \text{Wait for a worker to finish.}\)
2. Compute posterior \(GP\).
3. Draw a sample \(g \sim GP\).
4. Re-deploy worker at \(\arg\max g\).

Synchronous: \textit{synTS}

At any given time,
1. \(\{(x'_m, y'_m)\}_{m=1}^M \leftarrow \text{Wait for all workers to finish.}\)
2. Compute posterior \(GP\).
3. Draw \(M\) samples \(g_m \sim GP, \forall m\).
4. Re-deploy worker \(m\) at \(\arg\max g_m, \forall m\).
**Experiment:** Branin-2D  \( M = 4 \)

Evaluation time sampled from a uniform distribution

---

![Graph showing simulation results for Branin-2D experiment with different algorithms: synRand, synUCBPE, synHUCB, synTS.](image-url)
Experiment: Branin-2D

Evaluation time sampled from a uniform distribution

$M = 4$

$\mathbf{SR}'(T)$ vs Simulated time units ($T$)
**Experiment:** Branin-2D \[ M = 4 \]

Evaluation time sampled from a uniform distribution
Experiment: Hartmann-18D \[ M = 25 \]

Evaluation time sampled from an exponential distribution
Black-box Optimisation methods are used in several scientific and engineering applications.

Bayesian Optimisation: A method for black-box optimisation which uses Bayesian uncertainty estimates for $f$.

Some modern challenges

- Multi-fidelity optimisation
- Parallel evaluations
- and several more . . .
Summary

- Black-box Optimisation methods are used in several scientific and engineering applications.
- Bayesian Optimisation: A method for black-box optimisation which uses Bayesian uncertainty estimates for $f$.
- Some modern challenges
  - Multi-fidelity optimisation
  - Parallel evaluations
  - and several more . . .

Thank you.

Slides are up on my website:  www.cs.cmu.edu/~kkandasa