

Optimización Bayesiana Robusta al Ruido: Algoritmos y Funciones de Adquisición para Funciones Objetivo Estocásticas

1. Introducción a la Optimización Bayesiana y el Desafío del Ruido Estocástico

1.1. Panorama de la Optimización Bayesiana Estándar

La Optimización Bayesiana (OB) es una estrategia secuencial de diseño para la optimización global de funciones de caja negra cuya evaluación es costosa.¹ Opera construyendo un modelo probabilístico sustituto (o subrogado) de la función objetivo y utilizando una función de adquisición para seleccionar los puntos más prometedores para la siguiente evaluación.¹ Este enfoque es particularmente útil cuando las evaluaciones de la función son costosas en tiempo o recursos, ya que busca minimizar el número de evaluaciones necesarias para encontrar el óptimo.

El **modelo sustituto** es típicamente un Proceso Gaussiano (PG). Un PG proporciona una distribución posterior sobre la función objetivo, lo que incluye predicciones de la media y la varianza (incertidumbre) en cualquier punto del espacio de búsqueda.¹ El PG se define por una función de media y una función de covarianza (o kernel), la cual codifica creencias previas sobre la suavidad de la función.⁶ La flexibilidad de los PGs para modelar funciones complejas y su capacidad para cuantificar la incertidumbre los hacen ideales para la OB.

La **función de adquisición** es una función de utilidad que guía la búsqueda equilibrando la exploración (muestreo en regiones de alta incertidumbre) y la explotación (muestreo en regiones donde es probable encontrar una mejora).¹ Ejemplos comunes incluyen la Mejora Esperada (EI, por *Expected Improvement*), la Probabilidad de Mejora (PI, por *Probability of Improvement*) y el Límite Superior de Confianza (UCB, por *Upper Confidence Bound*).¹ La elección de la función de adquisición es crucial, ya que determina la eficiencia de la búsqueda.

El **proceso iterativo** de la OB generalmente sigue estos pasos:

1. Inicialización: Se evalúa la función objetivo en unos pocos puntos iniciales, seleccionados aleatoriamente o mediante métodos como el Muestreo de Hipercubo Latino.¹
2. Ajuste/Actualización del modelo sustituto PG: Con los datos recolectados, se ajusta o actualiza el modelo PG.¹
3. Optimización de la función de adquisición: Se optimiza la función de adquisición

para seleccionar el siguiente punto a evaluar.¹

4. Evaluación de la función objetivo: Se evalúa la función objetivo en el nuevo punto seleccionado.¹
5. Iteración: Se añaden los nuevos datos al conjunto de observaciones y se repiten los pasos 2-4 hasta que se agota un presupuesto de evaluaciones predefinido o se alcanza un criterio de convergencia.¹

Este marco establece la base de la OB estándar, fundamental para comprender cómo la presencia de ruido estocástico introduce complicaciones significativas.

1.2. El Problema: Impacto del Ruido Estocástico en las Evaluaciones de la Función Objetivo y el Rendimiento de la OB

El ruido estocástico implica que la evaluación de la función objetivo $f(x)$ en el mismo punto x múltiples veces puede producir resultados diferentes. Esto se modela comúnmente como $y(x)=f(x)+\epsilon$, donde $y(x)$ es la observación ruidosa, $f(x)$ es el valor verdadero (latente) de la función y ϵ es un término de ruido aleatorio, a menudo asumido como Gaussiano con media cero.⁸ Este tipo de ruido es prevalente en experimentos físicos, simulaciones con componentes estocásticos (por ejemplo, en la ciencia de materiales o la dinámica molecular ¹⁰), o en la evaluación de modelos de aprendizaje automático basada en divisiones aleatorias de datos para entrenamiento y validación.⁹

La presencia de ruido induce varios **desafíos para la OB** ⁹:

- **Modelos Sustitutos Engañosos:** El ruido puede corromper las observaciones, llevando a un ajuste inexacto del modelo sustituto PG. Un PG mal ajustado afecta la calidad de la guía proporcionada por la función de adquisición, lo que puede ralentizar la convergencia o dirigir la optimización hacia regiones subóptimas.⁹ La información engañosa afecta la calidad del modelo sustituto ajustado y los puntos propuestos, lo que podría tener un impacto negativo en el resultado general de la optimización.⁹
- **Error de Identificación:** Las evaluaciones ruidosas dificultan la evaluación precisa del verdadero rendimiento de los puntos de diseño. El punto "óptimo" final encontrado podría no ser realmente el óptimo verdadero de la función subyacente libre de ruido.⁹ Esto significa que incluso si el algoritmo converge, podría hacerlo a una solución que no es la mejor en ausencia de ruido.
- **Rendimiento Degradado:** Se sabe que el ruido degrada significativamente el rendimiento de la OB. Puede causar la pérdida de convergencia o una degradación sustancial del rendimiento, especialmente si el ruido no es Gaussiano o se comporta de manera adversaria.¹¹ Los métodos estándar de OB a

menudo asumen observaciones libres de ruido o ruido Gaussiano simple (homoscedástico), suposiciones que frecuentemente no se cumplen en problemas del mundo real.¹¹

Un aspecto fundamental que explica por qué el ruido es tan problemático radica en la **dualidad de la incertidumbre del PG en entornos ruidosos**. En la OB estándar (sin ruido), la varianza predictiva del PG representa principalmente la incertidumbre epistémica, es decir, la falta de conocimiento sobre la verdadera función en regiones no exploradas. Esta varianza es crucial para el componente de exploración de las funciones de adquisición. Sin embargo, cuando las observaciones son ruidosas, el modelo PG incorpora esta fuente de variabilidad en sus predicciones de varianza. Si un único término de varianza representa tanto la incertidumbre epistémica como la incertidumbre aleatoria (el ruido inherente en las observaciones), la función de adquisición puede malinterpretar una alta varianza debida al ruido como una alta incertidumbre epistémica.¹¹ Esto puede llevar a una exploración ineficiente, como, por ejemplo, muestrear repetidamente un punto muy ruidoso sin necesariamente aprender más sobre la forma de la función subyacente en otras áreas. Esta "ambigüedad"¹¹ es una razón fundamental por la que se necesitan estrategias especializadas para el manejo del ruido, que puedan, idealmente, desenredar estas dos fuentes de incertidumbre o tener en cuenta explícitamente la estructura del ruido.

2. Caracterización y Modelado del Ruido Estocástico en Funciones Objetivo

Para abordar eficazmente el ruido en la Optimización Bayesiana, es crucial primero caracterizar y modelar adecuadamente la naturaleza de dicho ruido.

2.1. Tipos de Ruido

El ruido estocástico en las evaluaciones de la función objetivo puede manifestarse de diversas formas, siendo las más comunes:

- **Ruido Homoscedástico:** En este caso, la varianza del ruido, denotada como $\tau^2(x)$, es constante en todo el espacio de entrada; es decir, $\tau^2(x) \equiv \tau^2$ para alguna constante $\tau^2 > 0$.⁹ Esta es una suposición simplificadora común en muchos modelos de OB, donde se asume que la magnitud de la aleatoriedad en las observaciones no depende del punto específico que se está evaluando.
- **Ruido Heteroscedástico:** Aquí, la varianza del ruido $\tau^2(x)$ varía con la entrada x .⁹ Este escenario es a menudo más realista en aplicaciones prácticas. Por ejemplo, algunas regiones del espacio de parámetros de un modelo de aprendizaje automático pueden producir evaluaciones de rendimiento inherentemente más

variables que otras.⁹ Las tareas de ajuste de hiperparámetros, por ejemplo, a menudo exhiben heteroscedasticidad y no estacionariedad.¹³ La heteroscedasticidad plantea un desafío mayor, ya que el modelo de OB no solo debe aprender la función objetivo subyacente, sino también cómo varía la fiabilidad de sus observaciones.

La correcta identificación y modelado del tipo de ruido presente es fundamental, ya que dicta la complejidad del modelo sustituto requerido y la idoneidad de las diferentes estrategias robustas de OB.

2.2. Modelado del Ruido dentro del Marco del Proceso Gaussiano

El marco del Proceso Gaussiano (PG) es inherentemente flexible y puede adaptarse para manejar observaciones ruidosas.

- **Modelo de Observación:** Se asume que las observaciones $y(x)$ son una manifestación ruidosa de una función latente verdadera $f(x)$, tal que $y(x)=f(x)+\epsilon$, donde ϵ es el término de ruido, comúnmente modelado como una variable aleatoria Gaussiana con media cero y varianza $\tau^2(x)$, es decir, $\epsilon \sim N(0, \tau^2(x))$.¹⁰ El PG tiene como objetivo modelar la función latente $f(x)$ a partir de las observaciones ruidosas $y(x)$.
- **Ruido Homoscedástico en PG:** Para el ruido homoscedástico, la matriz de covarianza del PG, K , se aumenta sumando $\tau^2 I$ a su diagonal, donde I es la matriz identidad y τ^2 es la varianza del ruido constante.⁸ Esta varianza del ruido τ^2 (a menudo denotada como α en las implementaciones de software³) se convierte en un hiperparámetro del modelo PG que debe ser aprendido de los datos o establecido a priori.
- **Procesos Gaussianos Heteroscedásticos (PGH):** Para manejar el ruido heteroscedástico, donde la varianza del ruido $\tau^2(x)$ depende de la entrada x , se utilizan los PGH. Estos modelos a menudo emplean un segundo PG para modelar el logaritmo de la varianza del ruido, $\log(\tau^2(x))$, u otras formas paramétricas para capturar esta dependencia de la entrada.⁹
 - Por ejemplo, el algoritmo HEBO (Heteroscedastic Evolutionary Bayesian Optimisation) utiliza transformaciones de entrada y salida junto con un PG heteroscedástico para manejar procesos de ruido complejos y la no estacionariedad de la función objetivo.¹³
 - El cálculo de las distribuciones posteriores exactas para los PGH puede ser un desafío, ya que dejan de ser Gaussianas multivariadas; sin embargo, están surgiendo análisis teóricos que abordan estas complejidades.²⁰
- **Kriging con Efecto Pepita (Nugget):** En geoestadística, el campo del que se originó el Kriging (un sinónimo de modelado con PG), se puede añadir un término

"pepita" (nugget) a la diagonal de la matriz de covarianza. Este término puede representar el error de medición (ruido) o la variabilidad a microescala de la función que no es capturada por el modelo de covarianza espacial.¹⁰ Añadir un efecto pepita ayuda a estabilizar el cálculo de la inversa de la matriz de covarianza, especialmente con datos ruidosos o puntos de entrada casi colineales, y permite que el modelo no interpole exactamente los puntos de datos ruidosos, suavizando así la superficie estimada.²³

El modelado preciso del ruido dentro del modelo sustituto es el primer paso crucial hacia una OB robusta. Si el modelo de ruido está mal especificado, la función de adquisición operará sobre información errónea, lo que inevitablemente degradará el rendimiento de la optimización.

La transición de modelos de ruido homoscedástico a heteroscedástico aumenta significativamente la complejidad del modelo sustituto PG. Mientras que el ruido homoscedástico típicamente añade un solo hiperparámetro (τ^2) al PG, los modelos heteroscedásticos como los PGH pueden introducir muchos más hiperparámetros o incluso modelos adicionales completos (por ejemplo, un PG separado para modelar $\log(\tau^2(x))$). Esta mayor complejidad requiere más datos para un aprendizaje efectivo y puede hacer que la optimización de los hiperparámetros del propio modelo sustituto sea más desafiante. Existe, por lo tanto, un compromiso inherente: un modelo de ruido más flexible puede capturar mejor la realidad de un proceso ruidoso complejo, pero podría ser más difícil de ajustar de manera fiable, especialmente con presupuestos de evaluación limitados. Los modelos de ruido demasiado complejos podrían sobreajustarse si no se regularizan cuidadosamente o si los datos son escasos. Esto también subraya la importancia del aprendizaje eficiente de hiperparámetros para estos sustitutos más complejos, como se destaca en HEBO, que "admite la optimización exacta del logaritmo de la verosimilitud marginal".¹⁸ La elección del modelo de ruido debe, en consecuencia, equilibrarse con los datos disponibles y los recursos computacionales.

3. Estrategias Centrales Algorítmicas y de Funciones de Adquisición para la Optimización Bayesiana Robusta al Ruido

Una vez que se ha establecido un modelo sustituto que intenta capturar la naturaleza de la función objetivo y su ruido asociado, el siguiente paso crucial es definir cómo seleccionar el próximo punto para evaluar. Aquí es donde las funciones de adquisición especializadas y las estrategias algorítmicas para entornos ruidosos juegan un papel fundamental.

3.A. Mejora Esperada (EI) y sus Variantes para Entornos Ruidosos

La Mejora Esperada (EI) es una de las funciones de adquisición más populares en la OB.

- **Mejora Esperada Estándar (EI):** Se define como $EI(x) = E[\max(0, f(x) - f^*)]$, donde f^* es el mejor valor observado actualmente.⁷ En entornos ruidosos, f^* se toma a menudo como la observación ruidosa máxima o el máximo de la media posterior del PG.²⁶
 - *Limitaciones con Ruido:*
 - La EI estándar a menudo ignora la incertidumbre en el incumbente f^* cuando este se deriva de observaciones ruidosas.²⁶ Una observación ruidosa alta podría no representar un punto verdaderamente bueno de la función latente.
 - La EI, derivada originalmente para entornos sin ruido, puede funcionar mal cuando se aplica ingenuamente a observaciones ruidosas, especialmente con altos niveles de ruido.¹⁷
 - Puede sufrir una convergencia lenta si f^* se basa en la media del PG en presencia de mucho ruido.¹⁷
 - Presenta patologías numéricas: los valores de EI pueden volverse extremadamente pequeños o numéricamente cero en grandes partes del dominio, lo que dificulta la optimización basada en gradientes de la propia función de adquisición.²⁸

Debido a estas limitaciones, se han propuesto varias variantes de EI para mejorar su robustez al ruido:

- **Mejora Esperada Aumentada (aEI):** Picheny et al. (2013b) encontraron que este método funcionaba bien en problemas de referencia con ruido.¹⁷ Una forma generalizada es la "función de adquisición aumentada por ruido"¹¹, que multiplica la EI estándar por un término que aumenta con la varianza predictiva $\sigma^2(x)$. La fórmula es $AaEI = AEI \cdot [1 + (\sigma^2(x)/\epsilon)p]$.¹¹
 - *Mecanismo:* Esta aumentación recompensa explícitamente la exploración de puntos donde el modelo es incierto *debido al ruido*, promoviendo el remuestreo en ubicaciones ruidosas para refinar las estimaciones.
- **Mejora Esperada Corregida (cEI) / EI Modificada:** Propuesta por Zhou et al. (2023) [⁹ (refiriéndose a arXiv:2310.05166v1), ²⁶].
 - *Mecanismo:* Aborda la limitación de la EI estándar al incorporar explícitamente la incertidumbre de la solución incumbente. Considera una función de mejora con un valor objetivo desconocido sobre el mejor punto que maximiza la media predictiva del PG (similar a la EI sin ruido) pero deriva una expresión

analítica para esta función de adquisición *en presencia de observaciones ruidosas*. Utiliza directamente la *varianza correcta* y la información de covarianza del modelo PG.²⁶

- **Beneficio:** Proporciona una buena generalidad tanto para entornos ruidosos como sin ruido y aspira a reemplazar la fórmula estándar de EI en los paquetes de software.²⁶ Puede superar a la EI estándar en entornos ruidosos.²⁶
- **Propiedades Teóricas:** Alcanza una tasa de convergencia sublineal en el límite del arrepentimiento acumulado bajo ruido de observación heteroscedástico.²⁷
- **Mejora Esperada-Costo (EIC):** Adapta la EI para un mejor rendimiento bajo el arrepentimiento acumulativo introduciendo un "costo de evaluación". Un punto se muestrea solo si su valor de EI excede este costo; de lo contrario, se remuestrea el mejor punto actual.²⁵
 - **Mecanismo para el Ruido:** El remuestreo del mejor punto actual ayuda a refinar su estimación si es ruidoso, en lugar de explorar siempre puntos nuevos, potencialmente igualmente ruidosos.
- **LogEI:** Una reformulación de EI (y variantes como EHVI) que calcula $\log(1+EI)$ o una transformación similar, para remediar las patologías numéricas donde los valores de EI se vuelven demasiado pequeños para una optimización efectiva basada en gradientes.²⁸

Estas variantes de EI muestran una progresión en la adaptación de esta idea fundamental a las complejidades introducidas por el ruido, desde aumentaciones simples hasta reformulaciones más fundamentales.

3.B. Gradiente de Conocimiento (KG) en Entornos Ruidosos y Estocásticos

El Gradiente de Conocimiento (KG, por *Knowledge Gradient*) es otra función de adquisición importante.

- **KG Estándar:** Tiene como objetivo maximizar la mejora esperada en el valor del óptimo si se realizara una evaluación más. Es una política Bayes-óptima de un solo paso bajo ciertas condiciones.³⁰
- **KG para Observaciones Ruidosas:** Puede adaptarse para entornos ruidosos. Picheny et al. (2013b) encontraron que KG funcionaba bien en benchmarks ruidosos.¹⁷ El cálculo puede ser intensivo, a menudo requiriendo aproximación por Monte Carlo.¹¹
- **KG para Programación Estocástica:** Wu et al. aplican OB con una función de adquisición basada en KG para resolver programas estocásticos de dos etapas no convexos que son costosos de evaluar, optimizando el rendimiento promedio.⁹ Esto implica optimizar conjuntamente variables de primera etapa (decisiones

"aquí y ahora") y de segunda etapa (decisiones "esperar y ver").

- **KG con Gradientes Ruidosos:** Puede acomodar observaciones de gradientes incompletas y ruidosas si están disponibles.³⁰

KG ofrece un enfoque basado en la teoría de la decisión para seleccionar puntos. Su aplicación a la programación estocástica aborda directamente la optimización de funciones cuyo valor "verdadero" podría ser una expectativa sobre resultados ruidosos.

3.C. Adaptaciones de la Probabilidad de Mejora (Pol) y del Límite de Confianza Inferior/Superior (LCB/UCB)

Pol y UCB/LCB son funciones de adquisición fundamentales que también se han adaptado para el ruido.

- **Probabilidad de Mejora (Pol):** $\text{Pol}(x) = P(f(x) > f^* + \xi)$, donde ξ es un parámetro de compensación.⁸ Es un enfoque más simple y codicioso.¹¹
 - *Para Optimización Robusta de Valor Objetivo con Ruido:* Hoffer et al. (2023) derivan Pol para encontrar una entrada x tal que la salida $E(x)$ esté cerca de un valor objetivo y^* , minimizando el error cuadrático esperado $(m(x) - y^*)^2 + \sigma^2(x)$ (donde $\sigma^2(x)$ es la varianza aleatoria). Pol se convierte en $P(\hat{E}(x) \leq E_{\min} - \zeta)$, donde $\hat{E}(x)$ es el error cuadrático esperado estimado y sigue una distribución χ^2 no central.³²
- **Límite Superior de Confianza (UCB):** $\text{UCB}(x) = \mu(x) + \kappa\sigma(x)$.¹ Equilibra la explotación (alto $\mu(x)$) y la exploración (alto $\sigma(x)$).
 - *Adaptaciones para el Ruido:* El término $\sigma(x)$ incorpora naturalmente el ruido si el modelo PG lo hace. GP-UCB está diseñado para el arrepentimiento acumulativo y tiene garantías teóricas.⁷
 - *GP-UCB+ y EXPLOIT+:* Para entornos sin ruido (pero relevante para entender la mecánica de UCB), estos algoritmos complementan UCB o la maximización de la media posterior con exploración aleatoria para mejorar la distancia de llenado y los límites de arrepentimiento.³³
- **Límite Inferior de Confianza (LCB):** $\text{LCB}(x) = \mu(x) - \kappa\sigma(x)$.¹¹ Se utiliza para la minimización.
 - *Para Optimización Robusta de Valor Objetivo con Ruido:* Hoffer et al. (2023) derivan LCB para minimizar el cuantil- q de $\hat{E}(x)$, convirtiéndose en $\sigma^2(x)F_{1,\lambda(x)}^{-1}(q) + \sigma^2(x)$.³²

La adaptación de UCB y Pol para objetivos ruidosos específicos (como la optimización de valor objetivo) demuestra cómo los principios básicos de la OB pueden

especializarse.

3.D. Funciones de Adquisición Basadas en Teoría de la Información

Estas funciones buscan maximizar la información obtenida sobre la ubicación del óptimo u otras propiedades de la función.

- **Principio:** Ejemplos incluyen la Búsqueda Predictiva de Entropía (PES, por *Predictive Entropy Search*) y la Búsqueda de Entropía de Máximo Valor (MES, por *Max-value Entropy Search*).⁴
- **Rendimiento:** A menudo superan empíricamente a otras funciones de adquisición.⁴
- **Manejo del Ruido:** Pueden adaptarse a escenarios complejos, incluidas las observaciones ruidosas. PES se menciona en ¹⁰ como una función de adquisición investigada para superficies de respuesta intrínsecamente ruidosas.
- **Aproximación:** Pueden ser intratables de calcular exactamente, requiriendo aproximaciones.⁴

Estas funciones ofrecen una forma más fundamentada de guiar la exploración, lo que potencialmente conduce a una búsqueda más robusta en paisajes complejos y ruidosos.

3.E. Enfoques Especializados para la Incertidumbre Aleatoria y Epistémica

Algunos métodos avanzados se centran en distinguir y manejar explícitamente diferentes tipos de incertidumbre.

- **Optimización Bayesiana Robusta de Valor Objetivo** ³²:
 - *Objetivo:* Encontrar la entrada x tal que la salida de la caja negra $f(x)$ esté cerca de un valor objetivo y^* en términos del error cuadrático esperado, considerando el ruido aleatorio $\sigma^2(x)$. El objetivo es minimizar $E(x) = (m(x) - y^*)^2 + \sigma^2(x)$.
 - *Mecanismo:* Utiliza un PG para modelar $m(x)$ (incertidumbre epistémica $\sigma^2(x)$). El error cuadrático esperado normalizado $e(x) = (E^*(x) - \sigma^2(x)) / \sigma^2(x)$ sigue una distribución χ^2 no central. Las funciones de adquisición (Pol, EI, LCB) se derivan basándose en esta distribución.
 - *Idea Clave:* Separa la incertidumbre aleatoria y epistémica para optimizar tanto la cercanía al objetivo como la baja varianza de la salida.
- **Reducción Dirigida de Varianza (TVR, por Targeted Variance Reduction)** ³⁶:
 - *Objetivo:* Optimización robusta de E , donde Θ son parámetros de ruido.
 - *Mecanismo:* Una función de adquisición conjunta sobre los parámetros de control x y de ruido Θ . $TVR(x, \theta) = VR_n(x, \theta) \cdot \Phi(\dots)$, donde VR_n es la reducción de

varianza al muestrear (x, θ) y Φ es un término que equilibra la explotación y la exploración relacionado con $\mu_n(x) - \mu_n(x_n^*)$.

- *Idea Clave*: Permite un compromiso entre "exploración-explotación-precisión". La precisión se refiere a la reducción de la varianza en el objetivo $g(x) = E\Theta$. La evaluación es en forma cerrada.

Estos métodos representan una comprensión sofisticada del ruido, yendo más allá de las "observaciones ruidosas" para optimizar objetivos que involucran explícitamente características del ruido.

3.F. Optimización Bayesiana Heteroscedástica

Cuando el ruido es heteroscedástico, se requieren enfoques más especializados.

- **Procesos Gaussianos Heteroscedásticos (PGH)**: Como se discutió en la sección 2.2, los PGH modelan $\tau^2(x)$ como dependiente de la entrada.⁹ Esto es fundamental para las funciones de adquisición en la OB heteroscedástica. HEBO, por ejemplo, utiliza un PGH.¹⁹
- **HEBO (Heteroscedastic Evolutionary Bayesian Optimisation)**⁹:
 - *Mecanismo*: Un optimizador integral que ganó el Desafío de Optimización de Caja Negra de NeurIPS 2020. Combina varias estrategias:
 - **PG Heteroscedástico**: Modela el ruido dependiente de la entrada.
 - **Transformación no lineal de Entradas y Salidas (Warping)**: Transforma el espacio de entrada y los valores de salida para manejar la no estacionariedad y hacer que los datos sean más adecuados para el modelado con PG (por ejemplo, más parecidos a una distribución Gaussiana).¹³ La transformación de salida ayuda a gestionar la heteroscedasticidad.
 - **Algoritmo Evolutivo para la Maximización de la Adquisición**: Utiliza estrategias evolutivas para optimizar la(s) función(es) de adquisición, lo que puede ser más robusto para paisajes de adquisición complejos y multimodales.¹⁸
 - **Ensamblajes de Adquisición Multi-objetivo**: Optimiza un frente de Pareto de soluciones basado en múltiples funciones de adquisición simultáneamente, lo que permite un consenso o compromiso en lugar de depender de una única heurística.¹³
 - *Beneficio*: Robusto a varios desafíos en tareas de ajuste de hiperparámetros del mundo real, incluyendo heteroscedasticidad y no estacionariedad.
- **Optimización Bayesiana Arborescente Heteroscedástica**³⁸: Utiliza un modelo basado en árboles para particionar el espacio de entrada, permitiendo diferentes modelos PG (potencialmente con diferentes características de ruido) en

diferentes regiones para manejar la no estacionariedad y la heteroscedasticidad.

Estos son enfoques a nivel de sistema que combinan modelado sustituto avanzado con estrategias de adquisición sofisticadas para abordar problemas difíciles y ruidosos donde las características del ruido son complejas.

La evolución de las funciones de adquisición muestra una clara progresión desde funciones simples (Pol, EI básica, UCB) hacia otras cada vez más complejas y especializadas (cEI, variantes de KG, TVR, ensambles multi-objetivo en HEBO). Inicialmente, las funciones de adquisición proporcionaban un buen rendimiento base pero tenían modos de fallo o limitaciones conocidas, especialmente con el ruido. Los investigadores identificaron deficiencias específicas, como el manejo de la incertidumbre del incumbente por parte de EI, el costo computacional de KG o la estrategia de exploración específica de UCB. Esto llevó a versiones "corregidas" o "aumentadas" que abordaban estos problemas específicos (por ejemplo, cEI corrigiendo el manejo del incumbente, aEI añadiendo un término de varianza de ruido). Además, se introdujeron nuevos paradigmas (basados en teoría de la información, teoría de la decisión como KG, separación explícita de incertidumbre como TVR) para proporcionar soluciones más fundamentales o robustas. Finalmente, marcos como HEBO combinan múltiples avances (PGH, transformaciones, optimización evolutiva de múltiples funciones de adquisición) en un único y potente optimizador. Este campo está en activa evolución, con una tendencia hacia estrategias de adquisición más fundamentadas, robustas y, a veces, más complejas. Sin embargo, esto también implica la necesidad de una evaluación comparativa cuidadosa para comprender las ventajas y desventajas.⁶

Se puede observar una distinción entre el manejo implícito y explícito del ruido en las funciones de adquisición. Algunas funciones manejan el ruido implícitamente a través de la media y la varianza predictivas del PG (que reflejan el ruido si el modelo PG incluye un término de ruido, por ejemplo, UCB estándar, EI estándar con f^* ruidosa). Otras manejan el ruido explícitamente en su formulación (por ejemplo, la derivación de cEI, el término de varianza de ruido de aEI, el componente de precisión de TVR, Pol/EI/LCB para la optimización de valor objetivo con $\sigma^2(x)$). El manejo implícito depende en gran medida de que el modelo PG capture y represente correctamente el ruido. Si el modelo de ruido del PG es básico (por ejemplo, homoscedástico cuando el ruido es heteroscedástico), la función de adquisición recibe señales engañosas. El manejo explícito del ruido en las funciones de adquisición intenta incorporar la conciencia del ruido directamente en el cálculo de la utilidad, haciendo que la función de adquisición sea más directamente robusta a la presencia y características del ruido. Métodos como TVR o las funciones de optimización de valor objetivo³² van más

allá al hacer que aspectos del ruido (varianza aleatoria, efecto de los parámetros de ruido) formen parte del propio objetivo de optimización. En general, el manejo explícito del ruido en las funciones de adquisición es más robusto, especialmente cuando las estructuras de ruido son complejas o cuando el objetivo de la optimización en sí mismo depende del ruido.

A continuación, se presenta una tabla que resume algunas funciones de adquisición clave para la OB robusta al ruido:

Tabla 1: Funciones de Adquisición Clave para la Optimización Bayesiana Robusta al Ruido

Nombre de la Función de Adquisición	Mecanismo Central para el Manejo del Ruido	Tipo de Ruido Asumido/Manejado	Referencia(s) Clave(s)
Mejora Esperada Aumentada (aEI)	Multiplica EI por un término de varianza de ruido para fomentar el remuestreo.	Estocástico general	11
Mejora Esperada Corregida (cEI)	Corrige la derivación de EI para incumbente ruidoso usando covarianza.	Heteroscedástico, Estocástico general	26
Gradiente de Conocimiento (KG) (estoc.)	Optimiza el valor esperado sobre resultados estocásticos.	Programación estocástica	9
PoI (para valor objetivo)	Maximiza la probabilidad de alcanzar el error objetivo considerando la varianza aleatoria.	Aleatorio (varianza $\sigma^2(x)$)	32
LCB (para valor objetivo)	Minimiza el cuantil del error objetivo considerando la	Aleatorio (varianza $\sigma^2(x)$)	32

	varianza aleatoria.		
Reducción Dirigida de Varianza (TVR)	Optimiza conjuntamente explotación, exploración y precisión sobre parámetros de ruido.	Ruido paramétrico, Incertidumbre de entrada	36
Ensamble Multi-objetivo de HEBO	Optimización Pareto de múltiples funciones de adquisición con PGH y transformaciones.	Heteroscedástico, No estacionario	13
Búsqueda Predictiva de Entropía (PES)	Maximiza la ganancia de información sobre la ubicación del óptimo.	Estocástico general (con modelo de ruido apropiado en GP)	10

Esta tabla proporciona una visión comparativa rápida, permitiendo al lector ver diferentes enfoques lado a lado y facilitando el seguimiento a través de las referencias clave.

4. Optimización Bayesiana Distribucionalmente Robusta (DRBO) para la Incertidumbre en las Características del Ruido

Más allá de modelar y mitigar un proceso de ruido conocido o asumido, la Optimización Bayesiana Distribucionalmente Robusta (DRBO) ofrece un paradigma para manejar la incertidumbre sobre las propias características del ruido o las distribuciones de variables contextuales que lo influyen.

4.1. Concepto de Optimización Distribucionalmente Robusta (DRO)

La Optimización Distribucionalmente Robusta (DRO) busca encontrar soluciones que funcionen bien bajo la peor distribución de probabilidad posible dentro de un "conjunto de ambigüedad" de distribuciones plausibles.⁴¹ En lugar de asumir una única distribución verdadera para los parámetros inciertos (que podrían incluir el proceso de ruido), la DRO considera un conjunto de distribuciones cercanas a una distribución nominal (por ejemplo, una empírica construida a partir de muestras) y optimiza para el peor caso dentro de ese conjunto. Esto proporciona una forma de protegerse contra la especificación incorrecta del modelo de incertidumbre.

4.2. Optimización Bayesiana Distribucionalmente Robusta (DRBO)

La DRBO aplica los principios de la DRO al marco de la Optimización Bayesiana. Su objetivo es lograr robustez frente a cambios en los datos (data shifts) o incertidumbre en variables contextuales, que pueden manifestarse como ruido o influir directamente en él.⁴¹ El objetivo es encontrar soluciones que sean robustas a las peores distribuciones de contexto dentro de un conjunto de ambigüedad definido.⁴³ Este marco es particularmente útil cuando la distribución de las variables contextuales (que podrían afectar el ruido o la función objetivo directamente) es desconocida, solo parcialmente conocida a través de muestras, o puede cambiar con el tiempo.⁴¹

4.3. Enfoques Específicos de DRBO

Se han propuesto varios enfoques para DRBO, que difieren principalmente en cómo definen el conjunto de ambigüedad:

- **DRBO basada en MMD:** Kirschner et al. (2020) introdujeron la DRBO con conjuntos de ambigüedad definidos por la Discrepancia Máxima Media (MMD, por *Maximum Mean Discrepancy*).⁴¹ La MMD es una métrica para medir la distancia entre distribuciones. Este enfoque es computacionalmente viable para espacios de contexto discretos y de baja cardinalidad.
 - AIRBO es un algoritmo que utiliza MMD para modelar entradas inciertas de distribuciones arbitrarias, potenciando el Proceso Gaussiano subyacente.⁴⁵
- **DRBO de Wasserstein (WDRBO):** Utiliza conjuntos de ambigüedad definidos como bolas en la distancia de Wasserstein.⁴² La distancia de Wasserstein (a menudo llamada distancia de movimiento de tierras o *earth mover's distance*) mide el costo mínimo para transformar una distribución en otra. Este enfoque permite modelar de manera flexible la incertidumbre en distribuciones de contexto continuas manteniendo la tratabilidad computacional y aspira a proporcionar tasas de arrepentimiento de vanguardia en entornos basados en datos.⁴³
- **DRBO basada en ϕ -divergencia:** Considera la robustez frente al cambio de datos utilizando ϕ -divergencias, que subsumen opciones populares como la divergencia de Kullback-Leibler (KL) y la Variación Total.⁴¹ Esto ofrece una teoría general para la robustez con diferentes métricas de divergencia.
- **Satisfacción Bayesiana Robusta (RoBOS):** Para la OB contextual, busca soluciones robustas a cambios distribucionales no especificados mientras se logra una utilidad por encima de un umbral deseado, sin un conocimiento preciso del conjunto de ambigüedad.⁴⁴

La DRBO ofrece una forma más abstracta y potencialmente más poderosa de lograr

robustez cuando las características del ruido en sí mismas son inciertas o están sujetas a cambios. Es un paso más allá de asumir un modelo de ruido fijo (aunque sea complejo).

Este paradigma de DRBO puede verse como un meta-marco para la robustez. En lugar de ser un algoritmo único, es un enfoque que puede emplear diferentes métricas de distancia (MMD, Wasserstein, ϕ -divergencias) para definir conjuntos de ambigüedad y, por lo tanto, diferentes nociones de robustez. Mientras que la OB robusta estándar a menudo asume un modelo de ruido específico (por ejemplo, Gaussiano, PG heteroscedástico), la DRBO reconoce que este modelo de ruido asumido (o la distribución de contexto que influye en el ruido) podría ser en sí mismo incierto o cambiante. Al optimizar para la peor distribución dentro de un conjunto de ambigüedad, la DRBO busca la robustez a un nivel superior: la robustez a la especificación incorrecta del modelo con respecto al ruido o a los factores contextuales influyentes. La elección de la métrica de distancia para el conjunto de ambigüedad (Wasserstein, MMD, etc.) refleja diferentes supuestos sobre lo que constituye un conjunto "plausible" de distribuciones alternativas. Por lo tanto, la DRBO proporciona una forma más fundamentada de manejar las "incógnitas desconocidas" con respecto al proceso de ruido o los factores contextuales influyentes, pasando de la simple mitigación del ruido a una verdadera robustez distribucional. Esto es particularmente relevante para aplicaciones donde las distribuciones de datos pueden derivar con el tiempo o donde el entorno operativo introduce variaciones imprevistas.

5. Consideraciones Prácticas y Temas Avanzados

La implementación exitosa de la Optimización Bayesiana robusta al ruido requiere considerar varios aspectos prácticos y estar al tanto de algunos temas avanzados y desafíos persistentes.

5.1. Estrategias de Replicación y Asignación de Presupuesto

- **Concepto:** Evaluar el mismo punto de diseño múltiples veces (replicación) puede reducir la varianza de su estimación de rendimiento, especialmente para funciones ruidosas.⁹ El promedio de múltiples observaciones ruidosas en un punto x proporciona una estimación más fiable de $f(x)$.
- **Desafío:** Decidir cuánto presupuesto dedicar a replicar puntos existentes frente a explorar nuevos puntos es un dilema central.⁹ Realizar muchas replicaciones en un punto inferior es un desperdicio de evaluaciones costosas; realizar muy pocas en un buen punto puede llevar a una evaluación falsa de su rendimiento debido al ruido.⁹

- **Orientación:** Van Beek et al.¹⁰ sugieren que para presupuestos experimentales pequeños (por ejemplo, menos de 15d, donde d es la dimensionalidad), pocas o ninguna replicación podrían ser mejores. Para presupuestos entre 15d y 40d, replicar las muestras iniciales dos veces es recomendable. Para presupuestos mayores, podrían ser necesarias más replications y un mayor número de muestras iniciales (por ejemplo, 5d para menos de 30d, y 10d o más para presupuestos mayores). El presupuesto disponible es un factor muy consecuente para seleccionar otras configuraciones del esquema de OB.¹⁰
- **Conexión con Funciones de Adquisición:** Algunas funciones de adquisición, como la Mejora Esperada Aumentada (aEI), pueden fomentar implícitamente la reevaluación de puntos ruidosos al valorar la reducción de la varianza predictiva.¹¹

5.2. Técnicas de Promediado de Ruido

Más allá de la simple replicación, existen enfoques más sofisticados para promediar o integrar el efecto del ruido.

- Latham et al. (mencionado en ¹¹) utilizaron un esquema de Monte Carlo para construir múltiples modelos sustitutos a partir de realizaciones de ruido generadas artificialmente. Luego, calcularon una Mejora Esperada promediada sobre estos modelos, lo que promueve el remuestreo para mejorar el componente de error predictivo del error de convergencia.
- El trabajo sobre Optimización Bayesiana Restringida con Experimentos Ruidosos ¹⁷ deriva una expresión para la Mejora Esperada integrando directamente sobre la posterior de la función de adquisición, evitando heurísticas y manejando el ruido de una manera más bayesiana. Utilizan una aproximación de cuasi-Monte Carlo para una optimización eficiente. Este enfoque es particularmente relevante cuando tanto la función objetivo como las restricciones se observan con ruido.

5.3. Desafíos en la OB Robusta al Ruido

A pesar de los avances, persisten varios desafíos:

- **Alta Dimensionalidad:** La OB, en general, puede tener un rendimiento deficiente en funciones objetivo de alta dimensión (empíricamente, más de 20 dimensiones ⁶). El ruido exacerba este problema, ya que distinguir la señal del ruido se vuelve aún más difícil en espacios vastos. El ajuste del modelo sustituto GP y la optimización de la función de adquisición se vuelven significativamente más desafiantes.⁶
- **Ruido No Gaussiano o Complejo:** La mayoría de los métodos asumen que el ruido sigue una distribución Gaussiana. Si el ruido es no Gaussiano, no sub-Gaussiano, o tiene dependencias complejas (por ejemplo, correlacionado en

el tiempo o en el espacio), los modelos de ruido estándar de PG y las funciones de adquisición derivadas pueden ser inadecuados.¹¹ El trabajo de Mahapatra et al.¹¹ aborda específicamente un desafiante proceso de ruido no Gaussiano y no sub-Gaussiano en el descubrimiento de materiales.

- **Costo Computacional:** Los modelos de ruido más sofisticados (PGH, componentes de DRBO) y las funciones de adquisición (KG, basadas en teoría de la información, basadas en Monte Carlo) pueden aumentar significativamente la sobrecarga computacional.³⁰ Se utilizan a veces PGs dispersos y otras aproximaciones para mitigar este costo, pero pueden introducir sus propias aproximaciones y errores.⁵

Estos desafíos prácticos conducen a un **trilema inherente entre el presupuesto de evaluación, la robustez deseada y la complejidad del método de OB empleado.**

Existe una tensión natural: presupuestos limitados⁹ restringen el número de evaluaciones, lo que dificulta el aprendizaje de modelos de ruido complejos o la posibilidad de realizar una replicación extensiva. Lograr una alta robustez, especialmente contra ruido complejo (heteroscedástico, no Gaussiano) o cambios distribucionales (DRBO), a menudo requiere modelos y funciones de adquisición más sofisticados, que a su vez consumen más datos y son computacionalmente más caros.¹¹ Los métodos más simples pueden ser más prácticos para presupuestos muy pequeños, pero pueden ofrecer menos robustez. Por lo tanto, los profesionales deben considerar cuidadosamente estas compensaciones. El "mejor" método de OB robusta no es absoluto, sino que depende en gran medida de las restricciones y características específicas del problema. Esto también apunta a la necesidad de métodos adaptativos que puedan ajustar su complejidad o estrategia en función de los datos observados y el presupuesto restante.

Además, muchas de las elecciones discutidas (elección del kernel, función de adquisición, estrategia de replicación, tipo de modelo de ruido) son en sí mismas elecciones de diseño o hiperparámetros del algoritmo de OB. Esto implica un **"bucle externo" de diseño de la OB**. La selección de un kernel apropiado (por ejemplo, el kernel Matérn resultó superior en¹⁰), la elección de la función de adquisición correcta según el ruido y el objetivo¹⁰, y la decisión sobre las estrategias de replicación según el presupuesto y los niveles de ruido⁹ son cruciales. Este bucle externo de meta-optimización o selección cuidadosa de los componentes de la OB a menudo se guía por heurísticas, experiencia previa o estudios de ablación (como en el desarrollo de HEBO¹⁸). Aunque la OB automatiza la optimización de la función de caja negra, configurar el algoritmo de OB para un rendimiento robusto en entornos ruidosos requiere una experiencia y comprensión significativas de estos componentes

interactuantes. Esto motiva la investigación en técnicas de OB más automatizadas (AutoBO) o algoritmos más adaptativos.

La siguiente tabla compara diferentes marcos para la OB robusta:

Tabla 2: Comparación de Marcos para la Optimización Bayesiana Robusta

Marco/Enfoque	Suposiciones Clave sobre Ruido/Objetivo	Cómo se Modela/Maneja el Ruido	Fortalezas Principales	Limitaciones/Desafíos	Ejemplo de Algoritmos/Técnicas
OB Estándar con AFs Conscientes del Ruido	El ruido es Gaussiano, potencialmente heteroscedástico.	PG con término de ruido, AFs especializadas (cEI, aEI).	Relativamente más simple de implementar si el ruido se comporta bien.	Puede tener dificultades con ruido muy complejo o no Gaussiano.	PG + cEI ²⁶
OB Heteroscedástica (ej. HEBO)	El ruido es heteroscedástico y la función puede ser no estacionaria.	PGH, transformaciones de entrada/salida, optimización evolutiva de AF multi-objetivo.	Maneja eficazmente estructuras de ruido complejas y no estacionariedad.	Puede ser computacionalmente intensivo, muchos componentes que ajustar.	HEBO ¹³
OB Robusta de Valor Objetivo	El objetivo es igualar un valor objetivo minimizando el error cuadrático, considerando el ruido aleatorio.	Varianza aleatoria explícita en el objetivo, AFs especializadas basadas en χ^2 no central.	Optimiza la estabilidad alrededor de un objetivo.	Requiere conocimiento/estimación de la función de varianza aleatoria.	Hoffer et al. (2023) ³²

Optimización Bayesiana Distribucionalmente Robusta (DRBO)	La propia distribución del ruido/contexto es incierta dentro de un conjunto de ambigüedad.	Optimización sobre la peor distribución en el conjunto de ambigüedad (MMD, Wasserstein)	Proporciona robustez contra la especificación incorrecta del modelo de ruido/contexto.	Computacionalmente muy costoso, definir el conjunto de ambigüedad puede ser difícil.	WDRBO ⁴³ , AIRBO ⁴⁵
---	--	---	--	--	---

Esta tabla proporciona una comparación de alto nivel, ayudando a comprender las diferentes filosofías y supuestos que subyacen a diversas metodologías de OB robusta y guiando la selección de un marco apropiado.

6. Conclusión y Direcciones Futuras de Investigación

6.1. Resumen de Enfoques Clave

La optimización de funciones objetivo costosas y afectadas por ruido estocástico es un desafío persistente en numerosas aplicaciones científicas y de ingeniería. La Optimización Bayesiana ha surgido como un marco poderoso para abordar estos problemas de manera eficiente en cuanto a muestras. Para lograr robustez frente al ruido, la investigación se ha centrado en varias estrategias interconectadas:

1. **Modificación de Modelos Sustitutos:** Principalmente a través de Procesos Gaussianos que pueden incorporar ruido homoscedástico o, de manera más avanzada, ruido heteroscedástico (PGH). Técnicas como la transformación de entradas y salidas (warping) mejoran aún más la capacidad del modelo para capturar funciones complejas y no estacionarias afectadas por el ruido.¹³
2. **Diseño de Funciones de Adquisición Conscientes del Ruido:** Se han desarrollado numerosas funciones de adquisición que van más allá de las heurísticas estándar. Estas incluyen variantes de la Mejora Esperada (como aEI y cEI ¹¹), el Gradiente de Conocimiento adaptado para entornos ruidosos o estocásticos ¹⁷, funciones basadas en teoría de la información como PES ¹⁰, y enfoques que separan y manejan explícitamente la incertidumbre aleatoria y epistémica, como la Reducción Dirigida de Varianza (TVR) ³⁶ y las funciones para la optimización robusta de valor objetivo.³²
3. **Marcos Integrales:** Soluciones como HEBO combinan múltiples componentes avanzados (PGH, transformaciones, optimización evolutiva de ensambles de adquisición multi-objetivo) para ofrecer un rendimiento robusto en tareas complejas.¹³

4. **Optimización Bayesiana Distribucionalmente Robusta (DRBO):** Este paradigma aborda la incertidumbre en la propia distribución del ruido o de las variables contextuales, optimizando para el peor caso dentro de un conjunto de ambigüedad definido por métricas como MMD o la distancia de Wasserstein.⁴¹

Un tema unificador en todos estos métodos avanzados es la **cuantificación y utilización sofisticada de diversas formas de incertidumbre**. La OB básica utiliza la varianza del PG para la exploración. La OB robusta al ruido necesita distinguir la varianza del ruido de la varianza del modelo. Los métodos avanzados como cEI utilizan explícitamente términos de covarianza que reflejan las relaciones de incertidumbre. La DRBO cuantifica la incertidumbre sobre la propia distribución del ruido. Los avances futuros en la OB robusta al ruido probablemente provendrán de formas aún más matizadas de modelar, desenredar y aprovechar estratégicamente estas múltiples facetas de la incertidumbre en el proceso de toma de decisiones (es decir, la función de adquisición).

6.2. Desafíos Abiertos y Posibles Vías para Trabajos Futuros

A pesar de los progresos significativos, la OB robusta al ruido sigue siendo un área de investigación activa con varios desafíos abiertos y oportunidades para futuras contribuciones:

- **Escalabilidad:** Existe una necesidad continua de métodos que escalen eficientemente a problemas de mayor dimensionalidad y a un mayor número de evaluaciones, manteniendo al mismo tiempo la robustez al ruido.⁴⁸ El "mal de la dimensionalidad" y el costo cúbico de los PGs estándar siguen siendo barreras.
- **Ruido Altamente Complejo:** Mejorar el manejo de ruido no Gaussiano, correlacionado (por ejemplo, en series temporales o datos espaciales), o incluso ruido adversario, donde la naturaleza del ruido podría cambiar en respuesta a las acciones del optimizador.¹¹
- **Optimización Bayesiana Automatizada para Entornos Ruidosos (AutoBO):** Desarrollar métodos que puedan seleccionar o adaptar automáticamente modelos de ruido, funciones de adquisición y otros hiperparámetros de la OB para un rendimiento robusto óptimo, reduciendo la carga sobre el usuario experto.
- **Comprensión Teórica:** Se necesita un mayor desarrollo de los límites de arrepentimiento y las garantías de convergencia para modelos de ruido más complejos y funciones de adquisición, especialmente bajo ruido heteroscedástico o no estándar, y para funciones objetivo que no residen necesariamente en un Espacio de Hilbert de Kernel Reproductor (RKHS).²⁶
- **Evaluación Comparativa (Benchmarking):** La comunidad necesita benchmarks

completos y estandarizados para la OB robusta al ruido que permitan una comparación justa y rigurosa de las técnicas emergentes.⁶ Esto incluye definir funciones de prueba realistas con diversos tipos y niveles de ruido.

- **Practicidad de DRBO:** Hacer que los métodos DRBO sean más tratables computacionalmente y proporcionar una guía más clara sobre la definición de conjuntos de ambigüedad para la incertidumbre práctica del ruido en aplicaciones del mundo real.⁴¹
- **Integración de la Experiencia Humana:** Para problemas ruidosos del mundo real, desarrollar formas más efectivas de incorporar el conocimiento del dominio humano o las preferencias en el bucle de la OB robusta, especialmente cuando las evaluaciones son subjetivas o cualitativas.¹⁶

Abordar estos desafíos no solo mejorará la aplicabilidad y la fiabilidad de la Optimización Bayesiana en presencia de ruido estocástico, sino que también ampliará su impacto en una gama aún más amplia de dominios científicos y de ingeniería.

7. Referencias

- 1 GeeksforGeeks. (Consultado en 2024). Bayesian Optimization in Machine Learning. GeeksforGeeks.
- 2 Pyro.ai. (Consultado en 2024). Bayesian Optimization. Pyro.ai Examples.
- 9 Moosbauer, J. (2021). Sequential Model-Based Optimization for Expensive Black-Box Functions with Heteroscedastic Noise. Ludwig-Maximilians-Universität München. (PDF)
- 11 Mahapatra, D. R., et al. (2022). An augmented Bayesian optimization approach for materials discovery with noisy and unreliable measurements. *Materials Advances*, 3(8), 3330-3344.
- 43 Papini, M., et al. (2024). Wasserstein Distributionally Robust Bayesian Optimization. arXiv:2503.20341v1.
- 32 Hoffer, J. G., Ranftl, S., & Geiger, B. C. (2023). Robust Bayesian Target Value Optimization. *Computers & Industrial Engineering*, 180, 109279. (Preprint arXiv:2301.04344).
- 4 Martinez-Cantin, R. (2024). Information-theoretic Bayesian Optimization: Survey and Tutorial. arXiv:2502.06789v1. (Resumen de 4)
- 10 van Beek, A., et al. (2025). Bayesian Optimization for Intrinsically Noisy Response Surfaces. ResearchGate. (Preprint, consultado el 4 de marzo de 2025)
- 5 Number Analytics. (2023). Exploring Bayesian Optimization: Advanced Machine Learning. Number Analytics Blog.
- 41 Nguyen, Q., et al. (2021). Distributionally Robust Bayesian Optimization. *Proceedings of the 38th International Conference on Machine Learning (ICML)*. (Resumen en OpenReview)
- 46 StackExchange. (2022). Why does Bayesian optimization perform poorly in more than 20 dimensions? Cross Validated.
- 31 Wu, J., et al. (2024). Bayesian Optimization for Non-Convex Two-Stage Stochastic Optimization Problems. arXiv:2408.17387v1.
- 26 Zhou, H., Ma, X., & Blaschko, M. B. (2023). A Corrected Expected Improvement Acquisition

Function Under Noisy Observations. arXiv:2310.05166v3. (Resumen de 26)

4 Martinez-Cantin, R. (2025). Information-theoretic Bayesian Optimization: Survey and Tutorial. arXiv:2502.06789v1.

19 OptunaHub. (Consultado en 2024). HEBO (Heteroscedastic and Evolutionary Bayesian Optimisation). OptunaHub Samplers.

20 Kim, S., et al. (2024). Theoretical Analysis of Heteroscedastic Gaussian Processes with Posterior Distributions. arXiv:2409.12622v1.

17 Letham, B., et al. (2019). Constrained Bayesian Optimization with Noisy Experiments. *Bayesian Analysis*, 14(2), 495-519.

8 Krasser, M. (2018). Bayesian optimization. Martin Krasser's Blog.

24 SMT Development Team. (Consultado en 2024). Kriging (KRG). Surrogate Modeling Toolbox (SMT) Documentation.

23 Ranjan, P., et al. (2011). On the Choice of Nugget in Kriging Modeling for Deterministic Computer Experiments. *Technometrics*, 53(4), 384-392. (Resumen en ResearchGate)

36 Liu, M., et al. (2024). Targeted Variance Reduction: Robust Bayesian Optimization of Black-Box Simulators with Noise Parameters. arXiv:2403.03816.

43 Papini, M., et al. (2024). Wasserstein Distributionally Robust Bayesian Optimization. arXiv:2503.20341v1. (Resumen de 43)

42 Mohajerin Esfahani, P., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2), 115-166. (Tutorial relacionado 42)

25 Zhang, J., et al. (2023). Expected Improvement-Cost: A Novel Algorithm for Cumulative Regret Minimization in Bayesian Optimization. *Journal of Machine Learning Research*, 24(301), 1-34.

6 Bartels, S., et al. (2024). POLI: A Benchmark for High-Dimensional Bayesian Optimization of Discrete Sequences. *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

39 OpenReview. (2024). Review of POLI: A Benchmark for High-Dimensional Bayesian Optimization of Discrete Sequences. OpenReview.

35 Martinez-Cantin, R. (2025). Information-theoretic Bayesian Optimization: Survey and Tutorial. arXiv:2502.06789. (PDF)

12 van Beek, A., et al. (2025). Bayesian Optimization for Intrinsically Noisy Response Surfaces. arXiv:2503.00327. (PDF, versión de arXiv de 10)

29 KU Leuven AI. (2025). A Corrected Expected Improvement Acquisition Function Under Noisy Observations / Using Causality for Enhanced Prediction of Web Traffic Time Series. Leuven.AI Events. (Referencia a 26)

27 arXiv. (2023). A Corrected Expected Improvement Acquisition Function Under Noisy Observations. arXiv:2310.05166. (Página de resumen de 26)

26 Zhou, H., Ma, X., & Blaschko, M. B. (2023). A Corrected Expected Improvement Acquisition Function Under Noisy Observations. arXiv:2310.05166v1. (HTML de 26)

13 Cowen-Rivers, A. I., et al. (2020). HEBO: Heteroscedastic Evolutionary Bayesian Optimisation. *NeurIPS 2020 Black-Box Optimisation Challenge*. (Resumen en ResearchGate)

14 Cowen-Rivers, A. I., et al. (2022). HEBO: An Empirical Study of Assumptions in Bayesian

Optimisation. Journal of Artificial Intelligence Research, 74, 1269-1349.

21 Wang, H., et al. (2024). Bridging Bayesian Optimization and Surrogate-Assisted Evolutionary Algorithm: A Comparative Study on Model Uncertainty. arXiv:2403.14413v2.

38 Assael, J. A. M., et al. (2015). Heteroscedastic Treed Bayesian Optimisation. arXiv:1410.7172v2.

48 OpenReview. (2024). Review of "Interactive Bayesian Optimization with Imprecise Human Feedback". OpenReview.

40 Bartels, S., et al. (2024). POLI: A Benchmark for High-Dimensional Bayesian Optimization of Discrete Sequences. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks. (PDF de 6)

47 NeurIPS. (2024). POLI: A Benchmark for High-Dimensional Bayesian Optimization of Discrete Sequences. NeurIPS 2024 Virtual Poster Session. (Referencia a 6)

3 Number Analytics. (2023). Comprehensive Guide: Practical Bayesian Optimization Implementation. Number Analytics Blog.

45 OpenReview. (2023). Review of "AIRBO: Arbitrary Input uncertainty Robust Bayesian Optimization". OpenReview.

19 OptunaHub. (Consultado en 2024). HEBO (Heteroscedastic and Evolutionary Bayesian Optimisation). OptunaHub Samplers. (API y descripción de 19)

15 Cowen-Rivers, A. I., et al. (2022). HEBO: An Empirical Study of Assumptions in Bayesian Optimisation. Journal of Artificial Intelligence Research, 74, 1269-1349. (PDF de 14)

49 Jaiswal, A., et al. (2024). Approximation-Aware Bayesian Optimization. Proceedings of the 41st International Conference on Machine Learning (ICML). (PDF)

48 OpenReview. (2024). Review of "Interactive Bayesian Optimization with Imprecise Human Feedback". OpenReview. (Similar a 48)

26 Zhou, H., Ma, X., & Blaschko, M. B. (2023). A Corrected Expected Improvement Acquisition Function Under Noisy Observations. arXiv:2310.05166v1. (HTML de 26)

4 Martinez-Cantin, R. (2025). Information-theoretic Bayesian Optimization: Survey and Tutorial. arXiv:2502.06789v1. (HTML de 4)

52 Daulton, S., et al. (2022). Robust Multi-Objective Bayesian Optimization Under Input Noise. Proceedings of the 39th International Conference on Machine Learning (ICML). (PDF)

17 Letham, B., et al. (2019). Constrained Bayesian Optimization with Noisy Experiments. Bayesian Analysis, 14(2), 495-519. (PDF de 17)

25 Zhang, J., et al. (2023). Expected Improvement-Cost: A Novel Algorithm for Cumulative Regret Minimization in Bayesian Optimization. Journal of Machine Learning Research, 24(301), 1-34. (PDF de 25)

27 arXiv. (2023). A Corrected Expected Improvement Acquisition Function Under Noisy Observations. arXiv:2310.05166. (Página de resumen de 26)

19 OptunaHub. (Consultado en 2024). HEBO (Heteroscedastic and Evolutionary Bayesian Optimisation). OptunaHub Samplers. (Similar a 19)

18 Cowen-Rivers, A. I., et al. (2020). HEBO: Heteroscedastic Evolutionary Bayesian Optimisation. NeurIPS 2020 Black-Box Optimisation Challenge, Technical Report. (PDF)

16 Zhang, Z., et al. (2024). Anchor-based Preferential Bayesian Optimization with Heteroscedastic Noise. arXiv:2405.14657v1.

- 53 RBC Borealis. (Consultado en 2024). Tutorial 8: Bayesian Optimization. RBC Borealis AI Research Blogs.
- 54 The Moonlight. (Consultado en 2024). Noise-Aware Bayesian Optimization Approach. The Moonlight IO.
- 51 Li, Z., et al. (2025). On the Convergence Rate of Constrained Expected Improvement for Bayesian Optimization. arXiv:2505.11323. (PDF)
- 28 Daulton, S., et al. (2023). Unexpected Improvements: A Unifying Reformulation of Analytic EI, EHVI, and their Noisy, Constrained, and Parallel Variants. arXiv:2310.20708v3.
- 30 Wu, J. (2017). Knowledge Gradient Methods for Bayesian Optimization. (Tesis Doctoral, Cornell University).
- 44 OpenReview. (2023). Review of "Robust Bayesian Satisficing for Contextual Bayesian Optimization with Unknown Distribution Shifts". OpenReview.
- 3 Number Analytics. (2023). Comprehensive Guide: Practical Bayesian Optimization Implementation. Number Analytics Blog. (Similar a 3)
- 7 Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems (NIPS), 25. (PDF)
- 33 Papenmeier, L., et al. (2024). Bayesian Optimization with Noise-Free Observations: Improved Regret Bounds and Algorithms. arXiv:2401.17037v1.
- Referencias de la Búsqueda Adicional (S_B#):
- 9 Resultados de búsqueda de Google Scholar sobre "Bayesian optimization algorithms for stochastic noise ArXiv", "acquisition functions for noisy Bayesian optimization review ArXiv", "heteroscedastic Gaussian process Bayesian optimization", "Bayesian optimization with noise averaging techniques", "Kriging with nugget for noisy Bayesian optimization", "robust acquisition functions Bayesian optimization noisy".
- 11 Resultados de la búsqueda en la URL de pubs.rsc.org sobre la función de adquisición aumentada por ruido.
- 32 Resultados de la búsqueda en la URL de arxiv.org/abs/2301.04344 sobre funciones de adquisición para la optimización robusta de valor objetivo.
- 10 Resultados de la búsqueda en la URL de researchgate.net sobre OB para superficies de respuesta intrínsecamente ruidosas.
- 26 (Inaccesible, contenido inferido de 26)
- 36 Resultados de la búsqueda en la URL de arxiv.org/pdf/2403.03816 sobre la Reducción Dirigida de Varianza (TVR).
- 55 (Inaccesible, contenido inferido de 26)
- 37 (Inaccesible, contenido inferido de 13)
- 37 (Inaccesible, contenido inferido de 13)
- 27 (Inaccesible, contenido inferido de 26)
- 22 Resultados de la búsqueda en la URL de jair.org sobre HEBO.

Works cited

1. Bayesian Optimization in Machine Learning | GeeksforGeeks, accessed May 25, 2025,

- <https://www.geeksforgeeks.org/bayesian-optimization-in-machine-learning/>
2. Bayesian Optimization — Pyro Tutorials 1.9.1 documentation, accessed May 25, 2025, <https://pyro.ai/examples/bo.html>
 3. A Comprehensive Guide to Practical Bayesian Optimization Implementation, accessed May 25, 2025, <https://www.numberanalytics.com/blog/comprehensive-guide-practical-bayesian-optimization-implementation>
 4. Information-theoretic Bayesian Optimization: Survey and Tutorial - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2502.06789v1>
 5. Exploring Bayesian Optimization in Advanced Machine Learning - Number Analytics, accessed May 25, 2025, <https://www.numberanalytics.com/blog/exploring-bayesian-optimization-advanced-machine-learning>
 6. A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences - NIPS, accessed May 25, 2025, https://proceedings.neurips.cc/paper_files/paper/2024/file/fe0007fcfd707673660ec0f9014bc48e-Paper-Datasets_and_Benchmarks_Track.pdf
 7. Practical Bayesian Optimization of Machine Learning Algorithms - Harvard DASH, accessed May 25, 2025, <https://dash.harvard.edu/server/api/core/bitstreams/7312037c-e923-6bd4-e053-0100007fdf3b/content>
 8. Bayesian optimization - Martin Krasser's Blog, accessed May 25, 2025, <http://krasserm.github.io/2018/03/21/bayesian-optimization/>
 9. epub.ub.uni-muenchen.de, accessed May 25, 2025, https://epub.ub.uni-muenchen.de/74080/1/Moosbauer_Julia.pdf
 10. (PDF) Bayesian Optimization for Intrinsically Noisy Response Surfaces, accessed May 25, 2025, https://www.researchgate.net/publication/389547900_Bayesian_Optimization_for_Intrinsically_Noisy_Response_Surfaces
 11. Bayesian optimization for material discovery processes with noise ..., accessed May 25, 2025, <https://pubs.rsc.org/en/content/articlehtml/2022/me/d1me00154j>
 12. Bayesian Optimization for Intrinsically Noisy Response Surfaces - arXiv, accessed May 25, 2025, <https://arxiv.org/pdf/2503.00327>
 13. (PDF) HEBO: Heteroscedastic Evolutionary Bayesian Optimisation - ResearchGate, accessed May 25, 2025, https://www.researchgate.net/publication/346701349_HEBO_Heteroscedastic_Evolutionary_Bayesian_Optimisation
 14. HEBO: An Empirical Study of Assumptions in Bayesian Optimisation - ResearchGate, accessed May 25, 2025, https://www.researchgate.net/publication/361937474_HEBO_An_Empirical_Study_of_Assumptions_in_Bayesian_Optimisation
 15. HEBO: Pushing The Limits of Sample-Efficient Hyperparameter Optimisation - Journal of Artificial Intelligence Research, accessed May 25, 2025, <https://www.jair.org/index.php/jair/article/download/13643/26827/31229>
 16. Heteroscedastic Preferential Bayesian Optimization with Informative Noise

- Distributions, accessed May 25, 2025, <https://arxiv.org/html/2405.14657v1>
17. Constrained Bayesian Optimization with Noisy Experiments - Project Euclid, accessed May 25, 2025, <https://projecteuclid.org/journals/bayesian-analysis/volume-14/issue-2/Constrained-Bayesian-Optimization-with-Noisy-Experiments/10.1214/18-BA1110.pdf>
 18. HEBO: Heteroscedastic Evolutionary Bayesian Optimisation - NET, accessed May 25, 2025, <https://valohaichirpprod.blob.core.windows.net/papers/huawei.pdf>
 19. HEBO (Heteroscedastic and Evolutionary Bayesian Optimisation) - OptunaHub, accessed May 25, 2025, <https://hub.optuna.org/samplers/hebo/>
 20. Theoretical Analysis of Heteroscedastic Gaussian Processes with Posterior Distributions, accessed May 25, 2025, <https://arxiv.org/html/2409.12622v1>
 21. Model Uncertainty in Evolutionary Optimization and Bayesian Optimization: A Comparative Analysis - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2403.14413v2>
 22. HEBO: Pushing The Limits of Sample-Efficient Hyper-parameter ..., accessed May 25, 2025, <https://www.jair.org/index.php/jair/article/view/13643>
 23. On the Choice of Nugget in Kriging Modeling for Deterministic Computer Experiments | Request PDF - ResearchGate, accessed May 25, 2025, https://www.researchgate.net/publication/263566739_On_the_Choice_of_Nugget_in_Kriging_Modeling_for_Deterministic_Computer_Experiments
 24. Kriging (KRG) — SMT 2.9.2 documentation - Surrogate Modeling Toolbox - Read the Docs, accessed May 25, 2025, https://smt.readthedocs.io/en/latest/_src_docs/surrogate_models/gpr/krg.html
 25. Adjusted Expected Improvement for Cumulative Regret Minimization in Noisy Bayesian Optimization, accessed May 25, 2025, <https://www.jmlr.org/papers/volume26/22-0523/22-0523.pdf>
 26. A Corrected Expected Improvement Acquisition Function Under Noisy Observations - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2310.05166v1>
 27. [2310.05166] A Corrected Expected Improvement Acquisition Function Under Noisy Observations - arXiv, accessed May 25, 2025, <https://arxiv.org/abs/2310.05166>
 28. Unexpected Improvements to Expected Improvement for Bayesian Optimization - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2310.20708v3>
 29. A Corrected Expected Improvement Acquisition Function Under Noisy Observations / Using Causality for Enhanced Prediction of Web Traffic Time Series - Leuven.AI, accessed May 25, 2025, <https://ai.kuleuven.be/events/a-corrected-expected-improvement-acquisition-function-under-noisy-observations-using-causality-for-enhanced-prediction-of-web-traffic-time-series>
 30. KNOWLEDGE GRADIENT METHODS FOR BAYESIAN OPTIMIZATION, accessed May 25, 2025, <https://people.orie.cornell.edu/jdai/thesis/JianWuThesis.pdf>
 31. Bayesian Optimization for Non-Convex Two-Stage Stochastic Optimization Problems - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2408.17387v1>
 32. arxiv.org, accessed May 25, 2025, <https://arxiv.org/abs/2301.04344>
 33. Bayesian Optimization with Noise-Free Observations: Improved Regret Bounds

- via Random Exploration - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2401.17037v1>
34. [2502.06789] Information-theoretic Bayesian Optimization: Survey and Tutorial - arXiv, accessed May 25, 2025, <https://arxiv.org/abs/2502.06789>
 35. Information-theoretic Bayesian Optimization: Survey and Tutorial - arXiv, accessed May 25, 2025, <https://arxiv.org/pdf/2502.06789>
 36. arxiv.org, accessed May 25, 2025, <https://arxiv.org/pdf/2403.03816>
 37. accessed December 31, 1969, https://proceedings.neurips.cc/paper_files/paper/2020/file/555d6702c950ecb729a966b8c9194768-Paper.pdf
 38. [1410.7172] Heteroscedastic Treed Bayesian Optimisation - arXiv, accessed May 25, 2025, <https://arxiv.org/abs/1410.7172>
 39. A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences, accessed May 25, 2025, [https://openreview.net/forum?id=RJHQAcbmpZ&referrer=%5Bthe%20profile%20of%20Simon%20Bartels%5D\(%2Fprofile%3Fid%3D~Simon_Bartels1\)](https://openreview.net/forum?id=RJHQAcbmpZ&referrer=%5Bthe%20profile%20of%20Simon%20Bartels%5D(%2Fprofile%3Fid%3D~Simon_Bartels1))
 40. A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences - NIPS, accessed May 25, 2025, https://papers.nips.cc/paper_files/paper/2024/file/fe0007fcfd707673660ec0f9014bc48e-Paper-Datasets_and_Benchmarks_Track.pdf
 41. Distributionally Robust Bayesian Optimization with φ -divergences | OpenReview, accessed May 25, 2025, <https://openreview.net/forum?id=bRIEwWd7Vy-eld=3ObZGrz9Tg>
 42. Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning - Delft Center for Systems and Control, accessed May 25, 2025, https://www.dsc.tudelft.nl/~mohajerin/Publications/journal/2019/DRO_tutorial.pdf
 43. Wasserstein Distributionally Robust Bayesian Optimization with Continuous Context - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2503.20341v1>
 44. Robust Bayesian Satisficing - OpenReview, accessed May 25, 2025, [https://openreview.net/forum?id=xINPCvgULc&referrer=%5Bthe%20profile%20of%20Cem%20Tekin%5D\(%2Fprofile%3Fid%3D~Cem_Tekin2\)](https://openreview.net/forum?id=xINPCvgULc&referrer=%5Bthe%20profile%20of%20Cem%20Tekin%5D(%2Fprofile%3Fid%3D~Cem_Tekin2))
 45. Efficient Robust Bayesian Optimization for Arbitrary Uncertain inputs - OpenReview, accessed May 25, 2025, <https://openreview.net/forum?id=uDV4IA0gZ6-eld=jO0ysMmmoH>
 46. Why does Bayesian Optimization perform poorly in more than 20 Dimensions?, accessed May 25, 2025, <https://stats.stackexchange.com/questions/564528/why-does-bayesian-optimization-perform-poorly-in-more-than-20-dimensions>
 47. A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences, accessed May 25, 2025, <https://neurips.cc/virtual/2024/poster/97688>
 48. Principled Bayesian Optimization in Collaboration with Human Experts - OpenReview, accessed May 25, 2025, <https://openreview.net/forum?id=IDn9SiKgLy-eld=VziDb8gqqH>
 49. Approximation-Aware Bayesian Optimization - NIPS, accessed May 25, 2025,

https://proceedings.neurips.cc/paper_files/paper/2024/file/257b9a6a0e3856735d0e624e38fb6803-Paper-Conference.pdf

50. On the convergence rate of noisy Bayesian Optimization with Expected Improvement - arXiv, accessed May 25, 2025, <https://arxiv.org/html/2501.09262v2>
51. convergence rates of constrained expected improvement - arXiv, accessed May 25, 2025, <https://arxiv.org/pdf/2505.11323>
52. Robust Multi-Objective Bayesian Optimization Under Input Noise - arXiv, accessed May 25, 2025, <https://arxiv.org/pdf/2202.07549>
53. Tutorial #8: Bayesian optimization - Research Blog - RBC Borealis, accessed May 25, 2025, <https://rbcborealis.com/research-blogs/tutorial-8-bayesian-optimization/>
54. [Literature Review] Noise-Aware Bayesian Optimization Approach for Capacity Planning of the Distributed Energy Resources in an Active Distribution Network - Moonlight, accessed May 25, 2025, <https://www.themoonlight.io/en/review/noise-aware-bayesian-optimization-approach-for-capacity-planning-of-the-distributed-energy-resources-in-an-active-distribution-network>
55. accessed December 31, 1969, <https://arxiv.org/pdf/2310.05166>