

# Optimisation bayésienne par méthodes SMC

Romain Benassi\*, Julien Bect, Emmanuel Vazquez

13 janvier 2012

## 1 Contexte

L'optimisation bayésienne consiste à utiliser un modèle probabiliste de la fonction objectif afin de guider au mieux la recherche de l'optimum. Cette approche peut s'avérer particulièrement pertinente lors de l'optimisation de fonctions coûteuses, et trouve ainsi de nombreuses applications dans l'industrie et l'ingénierie en général. Chaque nouveau point d'évaluation considéré est celui correspondant à l'optimum d'un critère d'échantillonnage construit à l'aide des résultats d'évaluation déjà disponibles.

## 2 Description de la communication

Le problème considéré est la recherche du maximum global d'une fonction  $f : \mathbb{X} \rightarrow \mathbb{R}$ , avec  $\mathbb{X} \subset \mathbb{R}^d$ , en utilisant le critère *Expected Improvement* ou EI (voir [1] et ses références). L'idée principale de l'optimisation à l'aide du critère EI est d'utiliser une approche bayésienne :  $f$  est vue comme une réalisation d'un processus aléatoire  $\xi$  défini sur  $\mathbb{R}^d$ . Le processus  $\xi$  est généralement considéré gaussien conditionnellement à un paramètre  $\theta \in \Theta \subseteq \mathbb{R}^s$ , caractérisant ses fonctions moyenne et de covariance. Par la suite, après s'être donné un *a priori*  $\pi_0$  sur  $\theta$  et quelques résultats d'évaluations  $\xi(X_1), \dots, \xi(X_{n_0})$  aux points  $X_1, \dots, X_{n_0}$ , l'algorithme construit une suite de points d'évaluation  $X_{n_0+1}, X_{n_0+2}, \dots$  tel que, pour tous  $n \geq n_0$

$$X_{n+1} = \operatorname{argmax}_{x \in \mathbb{X}} \bar{\rho}_n := \int_{\theta \in \Theta} \rho_n(x; \theta) d\pi_n(\theta), \quad (1)$$

avec  $\pi_n$  la distribution *a posteriori* de  $\theta$ , conditionnellement à la  $\sigma$ -algèbre  $\mathcal{F}_n$  générée par  $X_1, \xi(X_1), \dots, X_n, \xi(X_n)$ , et

$$\rho_n(x; \theta) := \mathbb{E}_{n, \theta}((\xi(X_{n+1}) - M_n)_+ | X_{n+1} = x)$$

valeur de l'EI au point  $x$  sachant  $\theta$  et  $\mathcal{F}_n$ , avec  $M_n = \xi(X_0) \vee \dots \vee \xi(X_n)$  et  $\mathbb{E}_{n, \theta}$  l'espérance conditionnelle à  $\mathcal{F}_n$  et  $\theta$ . Si  $\xi$  est choisi gaussien, le critère  $\rho_n$  peut s'écrire sous la forme d'une expression analytique permettant de le calculer de façon efficace. Cependant, deux problèmes concernant la complexité, ainsi que les mises en œuvre possibles, se posent. a) Comment calculer une approximation de l'intégrale (1) ? b) De quelle façon procéder pour optimiser  $\bar{\rho}_n$  à chaque étape ?

Notre travail [2] apporte une réponse conjointe à ces deux questions à l'aide d'une approche SMC (sequential Monte Carlo) [3, 4]. L'idée générale est de considérer un ensemble pondéré  $\mathfrak{T}_n = \{(\theta_{n,i}, w_{n,i}) \in \Theta \times \mathbb{R}, 1 \leq i \leq I\}$  échantillonné selon  $\pi_n$ , afin de calculer une approximation de  $\bar{\rho}_n$  de sorte que,  $\sum_{i=1}^I w_{n,i} \rho_n(x; \theta_{n,i}) \rightarrow_I \bar{\rho}_n(x)$ . De plus, on associe à chaque itération un ensemble de points candidats  $\{x_{n,i,j}, 1 \leq j \leq J\}$ . L'idée est de répartir cet ensemble de points, pour chaque valeur de  $\theta$ , tel que  $\max_{i,j} \bar{\rho}_n(x_{n,i,j}) \approx \max_x \bar{\rho}_n(x)$ . Concrètement, nous construisons un ensemble de particules pondérées à chaque itération  $n$

---

\*Détenteur d'un Master recherche (SISEA) et d'un diplôme d'ingénieur (Télécom Bretagne), doctorant en troisième année à Supélec (département SSE) sous l'encadrement des deux autres auteurs, enseignants-chercheurs à Supélec

$\mathfrak{G}_n = \{ (\gamma_{n,i,j}, w'_{n,i,j}), \gamma_{n,i,j} = (\theta_{n,i}, x_{n,i,j}) \in \Theta \times \mathbb{X}, w'_{n,i,j} \in \mathbb{R}, 1 \leq i \leq I, 1 \leq j \leq J \}$ ,  
tel que  $\sum_{i,j} w'_{n,i,j} \delta_{\gamma_{n,i,j}} \rightarrow_{I,J} \pi'_n$ , avec

$$d\pi'_n(\gamma) = \tilde{g}_n(x | \theta) d\lambda(x) d\pi_n(\theta), \quad x \in \mathbb{X}, \theta \in \Theta, \gamma = (\theta, x),$$

où  $\lambda$  correspond à la mesure de Lebesgue,  $\tilde{g}_n(x | \theta) = g_n(x | \theta)/c_n(\theta)$ ,  $g_n(x | \theta)$  caractérise l'intérêt d'une évaluation en  $x$  (sachant  $\theta$  et les observations précédentes), et  $c_n(\theta)$  est un terme de normalisation valant  $\int_{\mathbb{X}} g_n(x | \theta) dx$ .

À l'initialisation, on génère un échantillon pondéré  $\mathfrak{T}_{n_0} = \{(\theta_{n_0,i}, w_{n_0,i}), 1 \leq i \leq I\}$  à partir de la loi  $\pi_{n_0}$ , et on choisit une densité  $q_{n_0}$  sur  $\mathbb{X}$  (densité uniforme, par exemple). Ensuite, pour  $n \geq n_0$ , l'algorithme procède en quatre étapes,

*Étape 1 : démarginalisation* — Utiliser  $\mathfrak{T}_n$  et  $q_n$  pour construire un ensemble  $\mathfrak{G}_n$ , avec  $x_{n,i,j} \stackrel{\text{iid}}{\sim} q_n$ ,  $w'_{n,i,j} = w_{n,i} \frac{g_n(x_{n,i,j} | \theta_{n,i})}{q_n(x_{n,i,j}) c_{n,i}}$ , et  $c_{n,i} = \sum_{j'=1}^J \frac{g_n(x_{n,i,j'} | \theta_{n,i})}{q_n(x_{n,i,j'})}$ .

*Étape 2 : évaluation* — Évaluer  $\xi$  en  $X_{n+1} = \operatorname{argmax}_{i,j} \sum_{i'=1}^I w_{n,i'} \rho_n(x_{n,i,j}; \theta_{n,i'})$ .

*Étape 3 : pondération/échantillonnage/déplacement* — Construire  $\mathfrak{T}_{n+1}$  à partir de  $\mathfrak{T}_n$  : repondérer les  $\theta_{n,i}$ s avec  $w_{n+1,i} \propto \frac{\pi_{n+1}(\theta_{n,i})}{\pi_n(\theta_{n,i})} w_{n,i}$ , rééchantillonner, et déplacer les  $\theta_{n,i}$ s pour obtenir les  $\theta_{n+1,i}$ s à l'aide d'un noyau de Metropolis-Hastings indépendant.

*Étape 4 : construire  $q_{n+1}$*  — Construire une estimation  $q_{n+1}$  de la seconde loi marginale de  $\pi'_n$  à partir de  $\mathfrak{X}_n = \{(x_{n,i,j}, w'_{n,i,j}), 1 \leq i \leq I, 1 \leq j \leq J\}$ .

Les résultats obtenus à partir d'une comparaison de notre algorithme avec un algorithme de référence (paramètre  $\theta$  estimé par MLE sur un grand jeu de données, et recherche des points candidats sur un LHS fixé) sont particulièrement encourageants comme tendent à le montrer les Figures 1(a) et 1(b).

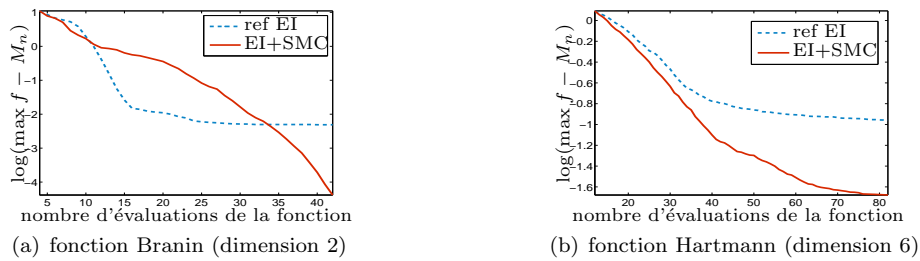


FIGURE 1 – Erreur moyenne à l'optimum sur une série de 100 tests

## Références

- [1] R. Benassi, J. Bect, and E. Vazquez. Robust Gaussian process-based global optimization using a fully Bayesian expected improvement criterion. In *LION5, proceedings*, Roma, Italy, 2011.
- [2] R. Benassi, J. Bect, and E. Vazquez. Bayesian optimization using sequential Monte Carlo. To appear in *LION6, proceedings*, Paris, France, 2012.
- [3] N. Chopin. A sequential particle filter method for static models, *Biometrika*, 89(3) :539–552, 2002.
- [4] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers, *J. R. Stat. Soc. B*, 68(3) :411–436, 2006.