Étude d'un nouveau critère d'optimisation bayésienne

Romain Benassi*

16 janvier 2010

1 Contexte

L'optimisation bayésienne consiste à utiliser un modèle probabiliste de la fonction à optimiser 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 (voir par exemple [8, 3]). Une des idées importantes est de choisir de façon itérative les points d'évaluation. Chaque nouveau point d'évaluation considéré est celui correspondant à l'optimum d'un critère d'échantillonage construit à l'aide des résultats d'évaluation déjà disponibles. Généralement les critères utilisés sont ceux permettant un bon compromis entre recherche locale et globale. Maximiser l'espérance de l'amélioration, ou l'EI pour *Expected Improvement* [6], est un critère bien connu et souvent utilisé. Cependant d'autres critères, éventuellement plus performants dans certaines situations, existent (voir l'algorithme IAGO dans [9]).

2 Algorithmes EI et EI à deux pas

Considérons f une fonction à maximiser, modélisée par un processus stochastique F, et $x_1, \ldots, x_n$ des points d'évaluation de f. Soit $M_n = \max\{F(x_i) | 1 \leq i \leq n\}$ le maximum courant. L'algorithme EI consiste à choisir un point d'évaluation x_{n+1} qui maximise

$$EI^n(x) = \mathbb{E}\Big\{I_n(x) | F(x_1), \ldots, F(x_n)\Big\},$$

avec

$$I_n(x) = \begin{cases} F(x) - M_n & \text{si } F(x) > M_n, \\ 0 & \text{sinon.} \end{cases}$$

Cette quantité correspond à l'amélioration moyenne, conditionnée par les évaluations précédentes, apportée par une évaluation en x. Ici, Le processus F est supposé gaussien, et la théorie du *krigeage* est utilisée pour prédire F au point x.

L'EI est un critère performant, qui permet un bon compromis entre recherche locale et globale [4]. L'algorithme EI est une stratégie bayésienne qui peut aussi être vue comme un problème de

*Master recherche (SISEA) et ingénieur Télécom Bretagne, doctorant à Supélec (département SSE) depuis Octobre 2009, sur le sujet *Etude théorique d'algorithme d'optimisation bayésienne*. La thèse est encadrée par Julien Bect et Emmanuel Vazquez, enseignants-chercheurs à Supélec.

programmation dynamique avec un horizon d'un pas. L'extension de l'algorithme EI au cas d'un horizon supérieur à un est une idée attrayante. Ce problème se formalise sous la forme d'un système d'équations (une pour chaque évaluation permise lors d'une itération) pouvant se résoudre de façon récurrente. Dans cette optique, nous considérons un critère EI à deux pas. Le principe de ce critère, comme son nom l'indique, n'est plus de choisir le point apportant la meilleure amélioration pour la présente itération, mais celui qui, une fois choisi, apporte la meilleure amélioration pour l'itération prochaine. D'un point de vue théorique, le problème de l'optimisation à deux pas, appliqué après n_0 évaluations, peut se formaliser à l'aide de deux équations récurrentes (n équations pour un problème à n pas)

$$u_2(x) = \max_{x'} \mathbb{E}\{ \, (F(x') - M_n)_+ \, | F(x_1), \ldots, F(x_{n_0}), F(x) \, \}$$

$$u_1 = \max_{x} \mathbb{E}\{u_2(x) \mid F(x_1), \ldots, F(x_{n_0})\}.$$

Chacune des deux équations détermine un nouveau point d'évaluation à l'endroit où le maximum de l'expression à droite du signe égal est atteint.

Pour le moment, il semble qu'aucune étude n'ait été réalisée concernant les performances de l'EI à deux pas, même si le principe général a déjà été évoqué dans [5, 1, 7, 2]. Une étude comparative entre une implémentation de ce critère et l'EI classique, en utilisant comme fonctions à optimiser des trajectoires aléatoires et des fonctions tests, est effectuée. L'objectif est ainsi de comparer, à partir d'un grand nombre de trajectoires, deux algorithmes utilisant chacun l'un des critères. Les distances au maximum, au maximiseur ainsi que les valeurs relatives de l'EI représentent un moyen d'effectuer une première comparaison, et seront ainsi présentées. Ces résultats permettent d'apporter des élements de réponse sur une éventuelle amélioration apportée par l'EI à deux pas.

Références

[1] B. Betrò. Bayesian methods in global optimization. *Journal of Global Optimization*, 1 :1–14, 1991.

[2] D. Ginsbouger and R. Le Riche. Towards gp-based optimization with finite time horizon. *hal-00424309, version 1*, 2009.

[3] D. Ginsbourger. *Métamodèles multiples pour l'approximation et l'optimisation de fonctions numériques multivariables.* PhD thesis, Ecole nationale supérieure des Mines de Saint-Etienne, 2009.

[4] D. R. Jones, M. Schonlau, and J. William. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13 :455–492, 1998.

[5] J. Mockus. *Bayesian approach to global optimization.* Kluwer Academic Publisher, 1989.

[6] J. Mockus, V. Tiesis, and A. Zilinskas. The application of bayesian methods for seeking the extremum. *Towards Global Optimization*, 2 :117–129, 1978.

[7] M. A. Osborne, R. Garnett, and S. J. Roberts. Gaussian processes for global optimization. *3rd International Conference on Learning and Intelligent Optimization (LION3)*, 2009.

[8] J. Villemonteix. *Optimisation de fonctions coûteuses.* PhD thesis, Université Paris-Sud XI, Faculté des Sciences d'Orsay, 2008.

[9] J. Villemonteix, E. Vazquez, M. Sidorkiewicz, and E. Walter. Global optimization of expensive-to-evaluate functions : an empirical comparison of two sampling criteria. *Journal of Global Optimization*, 43(2-3) :373–389, 2009.