

Bayesian Optimization for Full Waveform Inversion

Bruno G. Galuzzi, Riccardo Perego, Antonio Candelieri, Francesco Archetti

Abstract Full Waveform Inversion (FWI) is a computational method to estimate the physical features of Earth subsurface from seismic data, leading to the minimization of a misfit function between the observed data and the predicted ones, computed by solving the wave equation numerically. This function is usually multimodal, and any gradient-based method would likely get trapped in a local minimum, without a suitable starting point in the basin of attraction of the global minimum. The starting point of the gradient procedure can be provided by an exploratory stage performed by an algorithm incorporating random elements. In this paper, we show that Bayesian Optimization (BO) can offer an effective way to structure this exploration phase. The computational results on a 2D acoustic FWI benchmark problem show that BO can provide a starting point in the parameter space from which the gradient-based method converges to the global optimum.

Keywords. Bayesian Optimization. Full Waveform Inversion. Inversion problems.

Bruno G. Galuzzi

Department of Computer Science, Systems and Communications, University of Milano-Bicocca, viale Sarca 336, 20125, Milan, Italy.
e-mail: brunogaluzzi@gmail.com

Riccardo Perego

Department of Computer Science, Systems and Communications, University of Milano-Bicocca, viale Sarca 336, 20125, Milan, Italy.
e-mail: r.perego12@campus.unimib.it

Antonio Candelieri

Department of Computer Science, Systems and Communications, University of Milano-Bicocca, viale Sarca 336, 20125, Milan, Italy.
e-mail: antonio.candelieri@unimib.it

Francesco Archetti

Consorzio Milano-Ricerche, via Roberto Cozzi, 53, 20126, Milan, Italy
Department of Computer Science, Systems and Communications, University of Milano-Bicocca, viale Sarca 336, 20125, Milan, Italy.
e-mail: francesco.archetti@unimib.it

1 Bayesian Optimization

Bayesian optimization (BO) [11] is a suitable global optimization algorithm to find a global minimum x^* of a black-box, usually expensive-to-evaluate, objective function $f(x)$, where $x \in X \subset \mathbb{R}^d$ is a point in a d -dimensional bounded-box space X .

In BO the objective function is modelled as a realization of a stochastic process, typically a Gaussian Process (GP) on a probability space (Ω, Σ, P) . A GP, which defines a prior distribution over the function f , is completely specified by its mean $\mu(x) : X \rightarrow \mathbb{R}$ and a definite positive covariance function $k(x, x') : X^2 \rightarrow \mathbb{R}$,

$$f(x) \approx GP(\mu(x); k(x, x')) \quad (1)$$

The BO algorithm starts with an initial set of k points $\{x_i\}_{i=1}^k \in X$ and the associated observations $\{y_i\}_{i=1}^k$, with $y_i = f(x_i)$. At each iteration $t \in \{k+1, \dots, N\}$, the GP prior is updated using the Bayes rule, to obtain posterior distribution conditioned on the current training set $S_t = \{(x_i, y_i)\}_{i=1}^t$ containing the past evaluated points and observations. For any point $x \in X$, the posterior mean $\bar{\mu}_t(x)$ and the posterior variance $\bar{\sigma}_t^2(x)$ of the GP, conditioned on S_t , are known in closed-form:

$$\mu_t(x) = K(X_t, x)^t [K(X_t, X_t) + \lambda I]^{-1} Y_t \quad (2)$$

$$\sigma_t^2(x) = k(x, x) - K(X_t, x)^t [K(X_t, X_t) + \lambda I]^{-1} K(X_t, x) \quad (3)$$

where $K(X_t, X_t)$ is the $t \times t$ matrix whose $(ij)^{th}$ entry is $k(x_i, x_j)$, $K(X_t, x)$ (respectively Y_t) is the $t \times 1$ vector whose i^{th} entry is $k(x_i, x)$ (respectively y_i) and λ is the noise variance. A new point x_{t+1} is then selected and evaluated to provide an observation $y_{t+1} = f(x_{t+1})$. This new pair $\{(x_{t+1}, y_{t+1})\}$ is added to the current training set S_t , to define the training set for the next iteration $S_{t+1} = S_t \cup \{(x_{t+1}, y_{t+1})\}$.

The new point to evaluate is selected by solving an auxiliary optimization problem, typically of the form:

$$x_{t+1} = \arg \max U_t(x; S_t) \quad (4)$$

where U_t is an acquisition function to maximize. The rationale is that, because the optimization run-time or cost is dominated by the evaluation of the expensive objective function f , time and effort should be dedicated to choosing a promising point to evaluate, by solving the auxiliary problem. Solving this auxiliary problem does not involve the evaluation of the expensive objective function f , but only the posterior quantities of the GP and, thus, is considered cheap.

In this paper we focus on two of the most used acquisition function: the expected improvement (EI) [12], and the confidence bound (CB) [1] (lower/upper confidence bound, LCB/UCB, for minimization and maximization problems, respectively). EI is defined as follows:

$$EI_t(x; S_t) = (f^+ - \mu_t(x)) \Phi\left(\frac{f^+ - \mu_t(x)}{\sigma_t(x)}\right) + \sigma_t(x) \cdot N\left(\frac{f^+ - \mu_t(x)}{\sigma_t(x)}\right) \quad (5)$$

where, considering a minimization problem, $f^+ = \min_{x_i \in X_{1:t}} f(x_i)$ is the best value found after t evaluations (aka best seen), and $\Phi(\cdot)$ and $N(\cdot)$ are the normal cumulative distribution and the density probability function, respectively. The LCB formula is:

$$LCB_t(x; S_t) = \mu_t(x) - k\sigma_t(x) \quad (6)$$

where $k \geq 0$ is a parameter to manage the exploration/exploitation trade-off: a larger k drives exploration. In this paper we investigate the application of BO to seismic inversion problems comparing the EI and CB acquisition functions. An example of its application in the model inversion optimization problems, in the field of haemodynamic, is in [15].

2 Introduction to seismic inversion problems

The estimation of the geological properties of the subsurface can be obtained by means of seismic acquisition (Fig. 1, a)), in which artificially induced seismic waves, created by an impulse source, propagate through the subsurface. The receivers, distributed on the surface along a line ($2D$ seismic) detect the returning waves and measure the arrival times and the amplitudes of the waves at different distances, or offset. The seismic data are organized in a seismograms $d_{obs}(t, x_r)$, where $t \in [0, T]$ is the recording time and $\{x_r\}_{r=1}^{nr} \subset \mathbb{R}^2$ is the set of receivers locations. Fig. 1, b) shows a seismogram acquired during a marine seismic acquisition.

A seismic inversion problem [7], in case of acoustic approximation, consists of the estimation of an acoustic velocity subsurface model $v_p(\mathbf{x})$, that explains the events observed in a seismogram (reflections, refractions), and can be formulated as a minimization problem

$$v^* = \arg \min_{v \in V} F(v) \quad (7)$$

with V the set of possible acoustic velocity models, and $F(v) \geq 0$ a misfit function

$$F(v) = \|d_{obs} - d_{pred}(v)\|, \quad (8)$$

measuring the difference between the observed and the predicted seismograms $d_{pred}(v)$, computed by means of some seismic modelling algorithm.

In case of $2D$ acoustic approximation, the generation and propagation of seismic waves is modelled by the $2D$ acoustic wave equation [13]:

$$\ddot{p}(\mathbf{x}, t) - v(\mathbf{x})^2 \Delta p(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_0) s(t) \quad (9)$$

where $t \in [0, T]$ is the recording time, $\mathbf{x} = (x, z) \in D \subset \mathbb{R}^2$ is a bi-dimensional space domain, p is the acoustic pressure of the wave, \mathbf{x}_0 is the location of the source, and $s(t)$ is the seismic source. The predicted seismograms d_{pred} correspond to the restriction of the solution of the acoustic wave equation to the receivers locations.

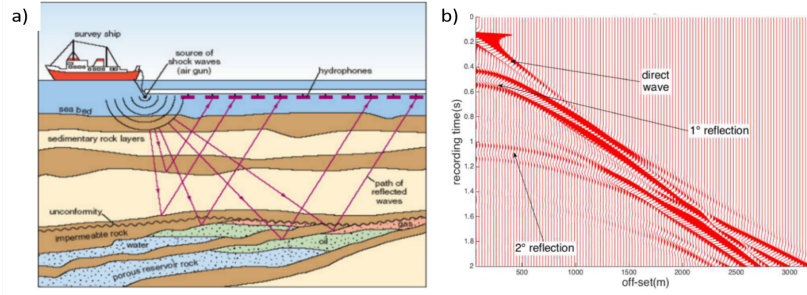


Fig. 1 a) Example of a marine seismic acquisition, taken from [19]. b) Example of a marine seismogram.

The solution of the wave equation can be obtained using an explicit finite-difference (FD) scheme, where the space domain D is sampled through a uniform grid spacing dx , along the horizontal and vertical direction, obtaining a regular grid D based on $nx \cdot nz$ grid nodes, with $i = 1, \dots, nx \cdot nz$, and with x_i , obtained scrolling the grid nodes along the rows.

It is important to remark that each evaluation of the misfit function is computationally expensive. Indeed, it requires the solution of the wave equation, which mandates for relatively small values of space sampling dx and time sampling dt [5]. According to a FD grid for the space parametrization, the acoustic FWI problem becomes an optimization problem with the number of variables given by the number of nodes in the modelling grid

$$\arg \min_{v \in V} F(v) \approx \arg \min_{v \in V^i} F(v^i) \quad (10)$$

where V^i represents the set of the P-wave velocity models discretized on the grid D . This means an optimization problem with $nx \cdot nz$ possible variables.

An important aspect of FWI is that the computation of the gradient of the misfit function, $\nabla_v f(v)$, can be done efficiently by means of the adjoint method [16]. This fact allows us to solve such minimization problem efficiently using an iterative procedure, updating an initial model v_0 with a gradient-based method, until a satisfactory match between the observed and the predicted data is obtained. However, as already noted in [2], the misfit functions are characterized by the presence of multiple local minima, and such local optimization approach will converge to a local minimum if the used starting model is not in the basin of attraction of the global minimum.

Global optimization algorithms can estimate such starting model, and meta-heuristic optimization methods have been already proposed for FWI, specifically Genetic Algorithms [17][10] and Simulated Annealing [6] [9]. In this paper we propose BO as global optimization approach to identify a promising starting model for a synthetic 2D acoustic FWI benchmark problem, in the field of seismic exploration.

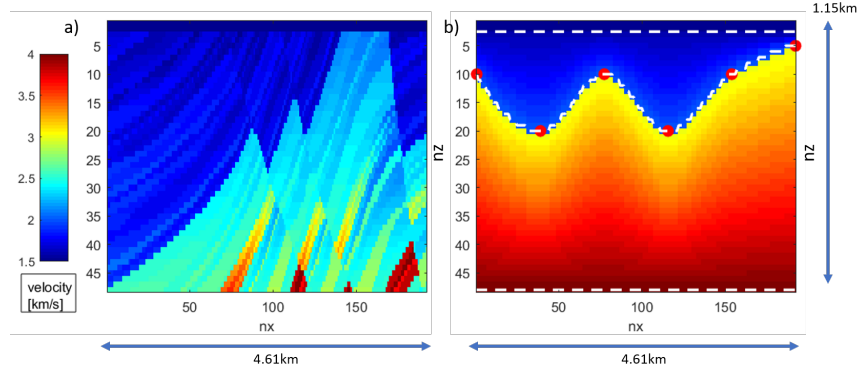


Fig. 2 a) The portion of the Marmousi velocity model used as true model for the FWI procedure. b) An example velocity model derived from the proposed sparse parametrization technique.

3 The Marmousi benchmark

This benchmark problem consists of the estimation of the acoustic velocity model of Fig. 2, a), that represents the upper central part of the Marmousi model [18], from a FWI procedure on a set of synthetic seismograms. This model contains 192 and 48 grid points in the z - and x -direction, respectively, with a grid spacing of 24 m. The first two rows of the modelling grid represent the water layer, whose velocity and depth are considered known and fixed a priori. We considered 16 seismograms, recorded by a spread of 192 receivers, equally spaced 24 m. Both the sources and the receivers are at a depth of 24 m, and the recording time is $T = 4$ secs. The synthetic seismograms are obtained by solving the acoustic wave equation (9), using an efficient FD scheme, whose details of the implementation can be found in [8]. As misfit function, we used the sum of the all the L^2 -norm difference between the observed and the synthetic seismograms, whereas as local optimization algorithm for FWI, we used the conjugate gradient method [14], one of most common gradient method to solve FWI [7].

4 Experimental setting and results

The starting model is estimated by means of BO using a sparse parameterization technique to reduce the number of parameters of modelling grid, formed by $192 \cdot 48$ grid nodes, to only 10 parameters, using a set of three interfaces and four velocities (Fig. 2, b)), as described in [6]. The first and the third interfaces are associated with the seabed (between 2nd and the 3rd row of the modelling grid) and the bottom of the model (situated at the 48th row) and are considered flat. The second interface represents a possible velocity contrast between the seabed and the bottom of the

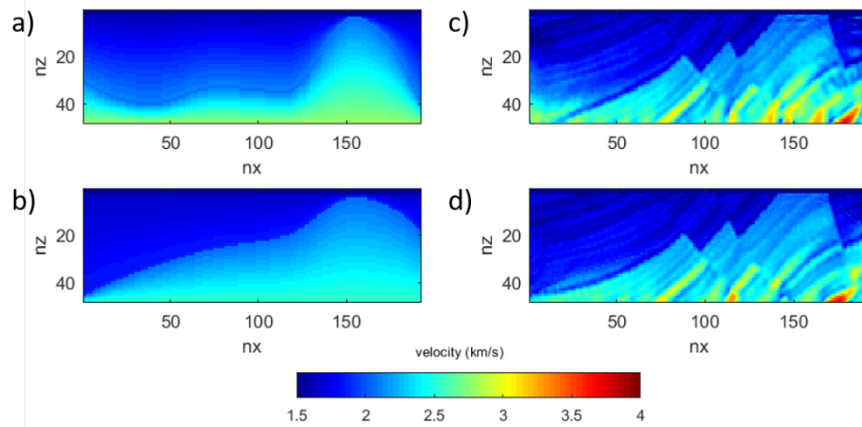


Fig. 3 a) Best velocity model obtained by EI. b) Best velocity model obtained by LCB. c) Final velocity model obtained by the local optimization using a) as starting model. d) Final velocity model obtained by the local optimization using b) as starting model.

model and is represented by a set of six nodes $\{(x_k, z_k)\}_{k=1}^6$ (the red point in Fig. 2, b)) with x_k evenly distributed and fixed along the x -direction and z_k that can range along the z -direction. We use a non-oscillatory spline to interpolate the interface across the nodes (the dot white spline of Fig. 2, b)). The four velocities prescribed the velocity just below the first interface, just above the second interface, just below the second interface, and just above the third interface, respectively. The velocities of the grid nodes between interfaces are obtained using linear interpolation in the vertical direction.

The overall number of variables needed for the parametrization is 10: a real variable for each one of the 4 velocities and 6 discrete variables for the z_k components of the nodes at the second interface. Velocities are in the range $1.5 - 4 \text{ km/sec}$, while the range for the 6 discrete variables is between 4 and 47. An overall number of 1000 function evaluations was fixed for each experiment, with 10 independent runs for each acquisition function (i.e., EI and LCB).

Fig. 3, a) and b) show the best velocity models obtained for EI and LCB, respectively. The first model obtains a value of the misfit function of 1868.1, whereas the second one obtains a value of 1723.5. These two models have been used as different starting points for the local optimization procedure on the modelling grid. Fig. 3, c) and d) show the corresponding velocity models obtained by the local optimization, after 1000 iterations of conjugate gradient. The final models obtained are very similar to the actual one, except in some areas near the lateral and the bottom boundaries, where, however, the seismic illumination is poor. The main result is that the starting models estimated by the BO, independently by the acquisition function used (EI vs LCB), can be considered quite near the basin of attraction of the global minimum of the misfit function, corresponding to the actual model. However, the differences between the two starting models obtained at the end of the BO, respectively for EI

(Fig 3.c) and LCB (Fig. 3,d)), lead to slightly different final models at the end of the entire procedure, with the lower misfit function value for LCB of 9.4 with respect to a value of 57.7 for EI.

5 Conclusions

According to the emerging interest on BO, to solve black-box and expensive optimization processes [3] [4], we have proposed the BO as a global optimization algorithm for the estimation of a promising starting model for FWI. We considered two alternative acquisition functions for BO and test them on a 2D acoustic FWI benchmark problem, namely the Marmousi model. The low error between the actual and the final models obtained, independently on the acquisition function, makes the proposed approach well suited for the seismic inversion problems, such as FWI, usually characterized by highly non-linearity, multiple local minima, and an expensive-to-evaluate misfit function.

References

1. Auer, P.: Using Confidence Bounds for Exploitation-Exploration Trade-off. *Journal of Machine Learning Research*. (2002) doi: 10.1162/153244303321897663
2. Beydon, W. B.: First Born and Rytov approximations: Modeling and inversion conditions in a canonical example. *The Journal of the Acoustical Society of America*. (1988) doi: 10.1121/1.396537
3. Candelieri, A., Giordani, I., Archetti, F., Barkalov, K., Meyerov, I., Polovinkin, A., Sysoyev, A., Zolotykh, N.: Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization. *Computers and Operations Research*. (2018) Article in Press.
4. Candelieri, A., Perego, R., Archetti, F.: Bayesian Optimization of pump operations in Water Distribution Systems. *Journal of Global Optimization*. Article in Press
5. Courant, R., Friedrichs K, Lewy, H.: On the Partial Difference Equations of Mathematical Physics. *IBM Journal*. (1967) doi: 10.1147/rd.112.0215
6. Datta, D., Sen, M.K.: Estimating a starting model for full-waveform inversion. *Geophysics*. (2016) doi: 10.1190/geo2015-0339.1
7. Fichtner, A.: Full seismic waveform modelling and inversion. *Springer Science & Business Media*. (2010) doi: 10.1007/978-3-642-15807-0
8. Galuzzi B., Zampieri, E., Stucchi, E.: A Local Adaptive Method for the Numerical Approximation in seismic wave modelling. *Communication in Applied and Industrial Mathematic*. (2017) doi: 10.1515/caim-2017-0014
9. Galuzzi B., Zampieri, E., Stucchi, E.: Global optimization procedure to estimate a starting velocity model for local Full Waveform Inversion. *Optimization and Decision Science: Methodologies and Applications*. (2017) doi: 10.1007/978-3-319-67308-0_18
10. Mazzotti, A., Bienati, N., Stucchi, E., Tognarelli, A., Aleardi, M., Sajeve, A.: Two-grid genetic algorithm full-waveform inversion. *The Leading Edge*. (2016) doi: 10.1190/tle35121068.1
11. Moćkus, J.: On Bayesian methods for seeking the extremum. *Optimization Techniques IFIP Technical Conference*. Springer. (1975) doi: 10.1007/978-3-662-38527-2_55

12. Moćkus, J.: Bayesian approach to global optimization: theory and applications. Springer Netherlands. (1989) doi: 10.1007/978-94-009-0909-0
13. Morton, K. W., Mayers D.F.: Numerical solution of partial differential equations: an introduction. Cambridge: Cambridge University Press. (2005)
14. Nocedal, J., Wright, S.: Numerical Optimization. Springer. (2006) doi: 10.1007/978-0-387-40065-5
15. Perdikaris P.,Karniadakis G.: Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. Journal of The Royal Society Interface. (2016) doi: 10.1098/rsif.2015.1107
16. Plessix, R.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophysical Journal International. (2006) doi: 10.1111/j.1365-246X.2006.02978.x
17. Sajeva, A., Aleardi, M., Stucchi, E., Bienati, N., Mazzotti, A.: Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model. Geophysics. (2016) doi: 10.1190/geo2015-0198.1
18. Versteeg, R.: The Marmousi experience: Velocity model determination on a synthetic complex data set. The Leading Edge. (1994) doi: 10.1190/1.1437051
19. "The Open University, Walton Hall, Milton Keynes, MK7 6AA", The Open University, Walton Hall, Milton Keynes, MK7 6AA, [Online]. Available: <http://www.open.edu/openlearn/>.