This is not the published version of the article / Þetta er ekki útgefna útgáfa greinarinnar

Author(s)/Höf.: Nosikov, I. A., Klimenko, M. V., Bessarab, P. F., & Zhbankov, G. A

Title/Titill: Application of the nudged elastic band method to the point-to-point radio wave ray tracing in IRI modeled ionosphere

Year/Útgáfuár: 2017

Version/Útgáfa: Pre-print / óritrýnt handrit

Please cite the original version:

Vinsamlega vísið til útgefnu greinarinnar:


Rights/Réttur: Published by Elsevier Ltd. All rights reserved.
Application of the nudged elastic band method to the point-to-point radio wave ray tracing in IRI modeled ionosphere

I.A. Nosikov\textsuperscript{a,b}, M.V. Klimenko\textsuperscript{a,b}, P.F. Bessarab\textsuperscript{c,d}, G.A Zbankov\textsuperscript{e}

\textsuperscript{a} Immanuel Kant Baltic Federal University, 236041, Kaliningrad, Russia, email: igor.nosikov@gmail.com

\textsuperscript{b} West Department of Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, RAS, 236017, Kaliningrad, Russia, e-mail: maksim.klimenko@mail.ru

\textsuperscript{c} Science Institute of the University of Iceland, 107 Reykjavik, Iceland, e-mail: bessarab@hi.is

\textsuperscript{d} ITMO University, St. Petersburg 197101, Russia

\textsuperscript{e} Southern Federal University, 344006, Rostov-on-Don, Russia, e-mail: gzhbankov@sfedu.ru

Correspondence to: I.A. Nosikov (igor.nosikov@gmail.com, phone: +74012215606; fax:+74012215606)

Abstract

Point-to-point ray tracing is an important problem in many fields of science. While direct variational methods where some trajectory is transformed to an optimal one are routinely used in calculations of pathways of seismic waves, chemical reactions, diffusion processes, etc., this approach is not widely known in ionospheric point-to-point ray tracing.
We apply the Nudged Elastic Band (NEB) method to a radio wave propagation problem. In the NEB method, a chain of points which gives a discrete representation of the radio wave ray is adjusted iteratively to an optimal configuration satisfying the Fermat’s principle, while the endpoints of the trajectory are kept fixed according to the boundary conditions. Transverse displacements define the radio ray trajectory, while springs between the points control their distribution along the ray. The method is applied to a study of point-to-point ionospheric ray tracing, where the propagation medium is obtained with the International Reference Ionosphere model taking into account traveling ionospheric disturbances. A 2-dimensional representation of the optical path functional is developed and used to gain insight into the fundamental difference between high and low rays. We conclude that high and low rays are minima and saddle points of the optical path functional, respectively.

*Keywords*: point-to-point ray tracing; ionospheric radio; Fermat’s principle; nudged elastic band method; IRI model; traveling ionospheric disturbances.

**Introduction**

Ray tracing problem essentially involves two steps. The first one is related to the choice of the environment model describing ionospheric parameters. Its accuracy has a direct impact on the agreement between modeled and experimental ionograms of oblique sounding. For quiet conditions, radio wave ray tracing calculations usually show good agreement with the observational data (Huang et al., 2006; Kotovich et al., 2010; Settimi et al., 2013), while storm time conditions cause problems. Another important issue concerns the methods for the ionospheric ray tracing, where position of the receiver and transmitter is fixed. The most traditional approach is the numerical solution of the eikonal equation combined with the shooting method also known as homing-in approach (Jones and Stephenson, 1975; Vasterberg, 1997; Coleman, 1998, 2011; Strangeways, 2000; Blagoveshchensky et al., 2009; Zhubkov et al., 2010; Karpachev et al., 2012). Recent calculations of radio waves propagating through the medium given by the IRI model and traveling ionospheric
disturbances (TID) reproduce well the observed Doppler frequencies and angles of arrival of ionospherically reflected high frequency (HF) waves (Huang et al., 2016). However, homing-in approach in the point-to-point ray tracing may suffer from convergence problems when applied to a realistic 3D ionosphere (Kalitkin, 1978). An alternative approach to this boundary-value problem is the use of variational methods based on direct minimization of radio ray optical path, where some initially defined trajectory is iteratively transformed to an optimal one while its end points are kept fixed according to the boundary conditions. This approach is widely used in seismology (Zhao et al, 1992; Koketsu and Sekine, 1998), where it is known as the bending method (Pereyra et al. 1980) and the pseudo bending method (Um and Thurber, 1987; Prothero et al. 1988; Moser et al., 1992), but hardly known in ionospheric radiophysics. Direct variational method for the point-to-point ionospheric ray tracing was proposed by Smilauer (1970) who derived ordinary differential equations for the radio ray and solved them using a Galerkin technique. Coleman (2011) developed an alternative approach involving discretization of the optical path functional. Direct variational method has advantages compared to the homing-in approach since it satisfies the boundary conditions automatically. However, low rays need a special treatment since they do not satisfy the Jacobi test for a minimum and, therefore, can not be found by direct minimization of the optical path functional (Coleman, 2011).

In this paper, the nudged elastic band (NEB) method, originally developed to identify minimum energy paths of chemical reactions (Jónsson et al., 1998), is applied to a point-to-point ionospheric ray tracing problem. High rays are calculated for the ionospheric medium predicted by International Reference Ionosphere (IRI) model (Bilitza (2001); Bilitza and Reinisch (2008); Bilitza et al. (2014)), where the electron density is either unperturbed or perturbed by travelling ionospheric disturbances (TIDs). The IRI model describes the climatology of ionospheric parameters (Eccles et al., 2011; Klimenko et al., 2015) and is widely used to calculate a medium for radio wave propagation. In this paper, the daytime summer solstice ionosphere obtained with IRI-2007 is used. Results obtained with the NEB method are compared with that given by homing-in approach (Zhbankov et al., 2010; Karpachev et al., 2012).

The problem of low ray identification by direct variational method is also discussed. For that reason a two-dimensional representation of the optical path surface is introduced and used to gain insight into the nature of low rays, which are particularly difficult to calculate, and to discuss a scheme for their identification.
Ray tracing method

In the isotropic medium, the optical path of the radio wave ray is defined by the following equation:

\[ S[\gamma] = \int_A^B n(r) \, dl \] (1)

Here, integration is performed along the curve \( \gamma \), which connects transmitter \( A \) and receiver \( B \); \( n(r) \) is the refractive index at point \( r \); \( dl \) is the length element along \( \gamma \). Continuous curve \( \gamma \) is then discretized into a contiguous sequence of linear segments so that the functional \( S[\gamma] \) becomes a multidimensional function of positions of \( N \) vertices defining a discrete representation of \( \gamma \). Ionospheric point-to-point ray tracing then reduces to an identification of stationary points of this function \( S(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) \), where \( \vec{r}_1, \vec{r}_2, ..., \vec{r}_N \) are positions on the vertices.

The antigradient of the optical path, \( -\nabla S = -\partial_i S \) which has a meaning of the force acting on the point in the multidimensional configuration space, can be used to guide the optimization. The use of the force in the optimization procedure, such as steepest descent method or conjugate gradient method, can lead to a problem connected with the discrete representation of the path \( \gamma \). Minimum of the optical path can correspond to a highly non-uniform distribution of the points, where there are several localization centers with very low density of points in between them. As a result, the information about the radio wave trajectory in some critical regions may be lost. The remedy to this problem lies in the force projection and inclusion of elastic forces, which was proposed in the Nudged elastic band (NEB) method (Mills and Jónsson, 1994). According to the NEB method, the force acting on each point \( i \) on the path \( \gamma \) is defined as

\[ \vec{F}_i = \vec{F}_i^t + \vec{F}_i^{\text{spring}} \] (2)

Here, \( \vec{F}_i^t = \frac{\partial S}{\partial \vec{r}_i} \left( \frac{\partial S}{\partial \vec{r}_i} \cdot \vec{r}_i \right) \cdot \vec{r}_i \) is a transverse component of \( -\nabla_i S \) while \( \vec{F}_i^{\text{spring}} = k \left( \vec{r}_i - \vec{r}_j \right) \cdot \vec{r}_i \) is the parallel component of the artificial spring force acting between the points, where \( \vec{r}_i \) is the unit vector tangent to the path at the point \( \vec{r}_i \), \( k \) is a spring constant. In a previously proposed method of transverse displacements, only projected forces \( \vec{F}_i^t \) were used to find radio wave trajectories (Nosikov et al., 2016), which reduced
computational costs but created a problem that control over the point distribution along the trajectory was lost. In the NEB method, $\vec{F}_1$ defines position of the trajectory in space, while $\vec{F}_{spring}$ controls the distribution of the points along the trajectory.

Some initial trajectory is needed to start an NEB calculation. This can be done in various ways, but the simplest method is to generate a linear interpolation between the endpoints. When two or more radio ray trajectories exist between the same receiver and transmitter, the optimization procedure will most likely lead to convergence to the trajectory closest to the initial path. In order to find all radio wave rays in such a situation, some sampling of the various trajectories needs to be carried out.

**Identification of high rays**

Point-to-point ionospheric ray tracing reduces to the identification of all high and low rays connecting the receiver and transmitter. Here we focus on the radio wave ray tracing between Kaliningrad and Stockholm at 8 MHz, where the electron density is given by IRI-2007 model for 12:00 UT on June 22, 2014 (see Figure 1). Results obtained with the homing-in approach (Zhbankov et al., 2010; Karpachev et al., 2012) are presented in Figure 2. Four radio wave rays have been found, two high and two low rays, which is consistent with a well-defined two-layer structure in electron density vertical profile. These solutions serve as a reference for the NEB calculations.

Both high rays can be calculated with the NEB method by setting initial guesses for the radio wave trajectory at the altitudes of F2 and E layer peaks (see Fig. 3). The results are in good agreement with the solutions given by the homing-in approach (see Fig. 2). However, the direct minimization method fails to converge on the low rays. This problem was discussed earlier by Coleman (2011). The low rays do not satisfy the Jacobi test for a minimum of the optical path functional and, therefore, can not be found by a direct minimization procedure.

**Analysis of low rays**

Low rays can still be found with the NEB method. For this, the whole trajectory of the radio wave needs to be divided at the apex and separate NEB calculations performed for each segment of the trajectory. The resulting trajectories coincide with the low rays obtained by the homing-in approach (see Fig. 4). However, apex position is usually unknown, which
reduces a predictive power of the NEB method for the low rays. In the next section, a two-dimensional, intuitive representation of the radio ray trajectory is developed and visualized on an optical path surface. This analysis gives deeper insight into the nature of the low rays and helps develop ways to calculate them efficiently.

**Optical path maps. Minima and saddle points**

The optical path given by the discretized functional (see Eq. (1)) is a function of many variables defining position of each vertex of the polygonal representation of the radio wave trajectory. In order to visualize this as a two-dimensional map, we use a reduced description of the model in terms of only two variables. This is accomplished by choosing a three-point representation of the radio wave trajectory, where two points are fixed according to the boundary conditions and the third one defines the apex position (hypothetical reflection point). Each segment of the trajectory is now a minimum of the optical path functional and is found by the NEB method. With this representation, the radio ray is completely defined by two variables – horizontal and vertical coordinates of the apex point – and a contour map of the optical path can be constructed.

Resulting contour map of the optical path is presented in Fig. 5, which demonstrates that high rays correspond to minima of the optical path, while the low rays correspond to saddle points. This explains why high rays can be reliably identified by direct minimization of the optical path. Saddle points are, however, difficult to locate. The difficulty arises from the need to minimize the optical path with respect to all but one degree of freedom for which a maximization should be carried out and it is not known a priori which degree of freedom should be treated differently. This problem can be solved with the Newton-Raphson method, as advocated by Coleman (2011). However, the Newton-Raphson method converges to any stationary point of an object function and does not discriminate between minima, maxima and saddle points of all orders, but our analysis suggests that the definite identification of the low rays is equivalent to the first order saddle point search, for which several methods have been developed, and the one which is very efficient and commonly used is actually the NEB method. Originally, the NEB method was introduced to calculate lowest-lying paths between minima of a multidimensional surface. A saddle point is extracted from the position of maxima along such paths. Therefore, the low ionospheric rays can be found by applying the NEB method in its original context. An optimal path needs to be found in a space of radio ray trajectories. The final, relaxed path obtained from an NEB calculation lies lowermost on the
multidimensional optical path surface so that the maximum along the path is precisely a saddle point corresponding to a low ray. Calculation of the low rays using this approach is a subject of future research.

**Effect of travelling ionospheric disturbances**

Travelling ionospheric disturbances (TIDs) have a strong impact on high frequency radio communication (Gershman and Grigorev, 1968; Oinats et al., 2016). The effect of TID on the electron density, $N$, is modeled by the following equation:

$$ N = N_0 \cdot (1 + \Delta N_1) \quad (3) $$

where $N_0$ is a background distribution, $\Delta N_1$ - is an irregularity. Background ionosphere is defined by the plasma frequency given by IRI-2007 model on a two dimensional grid. For a successful ray tracing, both plasma frequency and its gradient need to be continuous functions. This can be achieved by interpolating the grid data using cubic spline polynomials. As a result, an unperturbed electron density, $N_0(x,y,z)$, can be obtained at an arbitrary point.

Ionosphere irregularity and nonstationarity are modeled as a beam of several travelling monochromatic waves:

$$ N(r,t) = N_0(r,t) \cdot \left[ 1 + \sum_{i=1}^{M} \delta_i \sin \left( -\Omega_i t + \mathbf{P}_i \cdot \mathbf{r} + \Phi_{0i} \right) \right] \quad (4) $$

where $M$ is a number of TID harmonics, $N_0(r,t)$ is a space-time distribution of electron density from IRI-2007 modeled ionosphere; $\delta_i$ is a relative amplitude of the $i$th harmonics of the TID at frequency $\Omega_i$; $\mathbf{P}_i$ is a TID wave vector, $\Phi_{0i}$ is an initial phase of the $i$th harmonics.

Generally, each harmonics of the TID is determined by five parameters: relative amplitude, $\delta_i$; period, $T_i$; wave length, $\Lambda_i$; angle of inclination relative to the ground, $\Theta_i$; angle of propagation with respect to the ground, $\Psi_i$. Electron density distribution is then expressed in terms of these parameters

$$ N(r,t) = N_0(r,t) \cdot \left[ 1 + \sum_{i=1}^{M} \delta_i \sin \left( -\frac{2\pi}{T_i} t + \frac{2\pi}{\Lambda_i} x \cdot \cos \Theta_i \cdot \cos \Psi_i + \frac{2\pi}{\Lambda_i} y \cdot \cos \Theta_i \cdot \sin \Psi_i + \frac{2\pi}{\Lambda_i} z \cdot \sin \Theta_i + \Phi_{0i} \right) \right] \quad (5) $$
where \( \delta_i = \delta_{0i} \cdot \exp \left( \frac{z-Zm_i}{Lz_i} \right) \cdot \exp \left( \frac{x-Xm_i}{Lx_i} \right) \) is relative TID amplitude, \( \delta_{0i} \) is maximum relative TID amplitude; \( Xm_i, Zm_i \) are coordinates of TID maximum position [km]; \( Lx_i, Lz_i \) are TID scales [km]. Suppose TID propagates in a XZ-plane at \( t = 0 \). Finally formula (5) can be rewritten as

\[
N(r,t) = N_0(r,t) \cdot \left[ 1 + \sum_{i=1}^{N} \delta_{0i} \cdot \exp \left( \frac{z-Zm_i}{Lz_i} \right) \cdot \exp \left( \frac{x-Xm_i}{Lx_i} \right) \cdot \sin \left[ \frac{2\pi}{L_i} \cdot x \cdot \cos \Theta_i \cos \Psi_i + \frac{2\pi}{L_i} \cdot z \cdot \sin \Theta_i + \Phi_{0i} \right] \right]
\]

(6)

TIDs are assumed to propagate between high and mid latitude stations.

Fig. 5 shows the results of NEB calculations of all high and low rays (technique of fixing apex from solutions of homing-in approach) at 9 MHz between Kaliningrad and Tromsö for the midday summer conditions during medium solar activity (12:00 UT 22.06.2014), where the electron density given by IRI-2007 model is either unperturbed (fig. 5a) or perturbed by TIDs (fig. 5b). Simulation of TID propagation was given by setting the parameters: \( M = 1, \delta_{0i} = 0.8, Xm_i = 600 \text{ km}, Zm_i = 250 \text{ km}, Lx_i = 1600 \text{ km}, Lz_i = 100 \text{ km}, \Lambda_i = 200 \text{ km}, \Theta_i = 30^\circ, \Phi_{0i} = 30^\circ \). The presence of ionospheric disturbances results in a significant change in the ray trajectories. The high ray reflected from F2 layer refracts several times due to TIDs acquiring a complex, wave-like shape, which demonstrates the ability of the NEB method to find such radio wave trajectories. Another couple of high and low rays reflected from F1 layer and F2 layer, respectively, disappear when TIDs are present.

**Summary**

In this paper, we applied the NEB method to a point-to-point ionospheric ray tracing problem. Although the method was originally developed for calculations of mechanisms and pathways of chemical reactions, it proves to be well-suited for the identification of radio ray trajectories in realistic ionospheric media, especially when position of the receiver and transmitter are fixed. All high rays can be found given that some sampling of the initial conditions for the radio wave trajectory is performed. We used IRI-2007 model to simulate ionosphere, where the electron density is either unperturbed or perturbed by TIDs, and calculated radio rays between Kaliningrad and Stockholm as well as between Kaliningrad and Tromsö.
A care needs to be taken when calculating the low rays. Although both high and low rays are stationary radio wave trajectories, our analysis show that the former correspond to the minima of the optical path functional, while the latter correspond to the saddle points, which are difficult to locate. The low rays can still be found with the proposed technique if the trajectory is divided at the apex and separate calculations are performed for each segment of the radio ray. This scheme is, however, only possible if position of the apex is known. A better strategy is to exploit the NEB method, again, but in a different context. The NEB method was originally designed to identify saddle points on a multidimensional surface. In order to locate a low ionospheric ray, the NEB method needs to be applied to a path in a space of radio ray trajectories. The final, relaxed path lies lowermost on the multidimensional optical path surface so that the maximum along the path is precisely a saddle point corresponding to a low ray. Formulation of new methods for finding low rays can be based on conclusions drawn in the present work, and will be addressed in a future study.

While the effect of magnetic field has not been included in the present study, the generalization of the technique for radio rays propagating through magneto-active plasma should be straightforward. In this case, the optical path functional has a more complex form, where an integrand depends explicitly on the angle between the radio ray and Earth magnetic field (see, for example, Coleman (2011)). This angle can be readily accessed through a tangent to the radio wave path, which is already used in the method to project forces. Extension of the method to the propagation through magneto-active plasma is the subject of a future research.

The authors thank F.S. Bessarab and V.V. Klimenko for fruitful discussions. This study was financially supported by Grants from the RFBR No.15-35-20364, 16-35-00590, Russian target Federal program 5-100 and the task of the Ministry of Education and Science, project № 3.1127.2014/K.

References


Figure captions (print version)

Figure 1. Vertical profile of the plasma frequency above Kaliningrad (solid line) and Stockholm (dashed line) for daytime summer solstice 22.06.2014 obtained using IRI-2007.

Figure 2. Results of point-to-point ray tracing calculations using the homing-in method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014.

Figure 3. Results of point-to-point ray tracing calculations (only high rays) using the NEB method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014. Black dots connected with dashed lines represent an initial approximation for the rays; small grey dots show intermediate configurations of the trajectory during the iterative procedure.

Figure 4. Results of point-to-point ray tracing calculations (only low rays) using the NEB method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014. Black dots connected with dashed lines represent an initial approximation for the rays; small grey dots show intermediate configurations of the trajectory during the iterative procedure.

Figure 5. Optical path distribution for the frequency of 8 MHz in the ionosphere, where the electron density is given by the IRI model for 12 UT 22.06.2014. High and low rays (obtained with the NEB method) are shown with solid white lines.

Figure 6. Results of point-to-point ray tracing calculations (white solid lines) using the NEB method at frequency 9 MHz between Kaliningrad (54.57º N, 20º E) and Tromsø (65.65º N, 18.57º E) for daytime summer solstice on 22.06.2014. The electron density given by IRI-2007 model is either unperturbed (a) or perturbed by TIDs (b).
Figure captions (online version)

Figure 1. Vertical profile of the plasma frequency above Kaliningrad (blue line) and Stockholm (red line) for daytime summer solstice 22.06.2014 obtained using IRI-2007.

Figure 2. Results of point-to-point ray tracing calculations using the homing-in method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014.

Figure 3. Results of point-to-point ray tracing calculations (only high rays) using the NEB method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014. Black dots connected with dashed lines represent an initial approximation for the rays; small grey dots show intermediate configurations of the trajectory during the iterative procedure.

Figure 4. Results of point-to-point ray tracing calculations (only low rays) using the NEB method at frequency 8 MHz between Kaliningrad and Stockholm for daytime summer solstice on 22.06.2014. Black dots connected with dashed lines represent an initial approximation for the rays; small grey dots show intermediate configurations of the trajectory during the iterative procedure.

Figure 5. Optical path distribution for the frequency of 8 MHz in the ionosphere, where the electron density is given by the IRI model for 12 UT 22.06.2014. High and low rays (obtained with the NEB method) are shown with solid white lines.

Figure 6. Results of point-to-point ray tracing calculations (white solid lines) using the NEB method at frequency 9 MHz between Kaliningrad (54.57° N, 20° E) and Tromsø (65.65° N, 18.57° E) for daytime summer solstice on 22.06.2014. The electron density given by IRI-2007 model is either unperturbed (a) or perturbed by TIDs (b).