Description of `IGUG`: A MATLAB program for 3-D inversion of gravity data using graph theory

Saeed Vatankhah, Vahid Ebrahimzadeh Ardestani, Susan Soodmand Niri, Rosemary Anne Renaut and Hojjat Kabirzadeh

February 21, 2019(Update 2)

Abstract

This short report describes the open source MATLAB package `IGUG` for 3-D inversion of gravity data. The package contains a number of scripts and functions. The main algorithm implemented in `IGUG.m` is based on methodology that was introduced by [Bijani et al.(2015)], and the further developments that are described in [Vatankhah et al. (2018)]. To run `IGUG.m` it is assumed that the user has first used the script `GMD.m` to provide necessary data on the data set and the survey volume. `GMD.m` can be used to define a complete simulation or in conjunction with a real data set. For the real data set it is assumed that measured data and parameters on the survey volume are available. If the user has information about the noise distribution this can be provided, otherwise the user can indicate estimates used to find an approximation for a Gaussian noise distribution. For the simulation the user should specific the subsurface structure that generates the gravity anomaly. `IGUG.m` implements the genetic algorithm based on parameters selected through thhe `OptimParameters` dialogues. `FindtheSolution` determines the point mass distribution for the selected parameters and can be repeated for a given configuration for multiple choices of a regularization parameter. Script `Diagnostic_Results` is provided for analyzing and illustrating the results. Examples showing the use of the code are given, which were generated using the `publish` capability of MATLAB. Publishing the script for a given set of parameters will automatically produce an equivalent output for the indicated data set. The package also includes data files that can be used to reproduce the results presented on this web site. The package was developed and tested on MATLAB Version: 9.4.0.813654 (R2018a).

1 General Description

The package `IGUG` consists of several scripts and functions. These are described in the following sections, specifically addressing the main scripts for the generation of the data, the data inversion and the analysis of the results. There are a few general comments concerning the package.

Path The user should assure that the path is set to use the files that are in the directories private files and open files, keeping the main files `IGUG.m` and `GMD.m` in a higher level directory.

Dialogue Boxes In any question in which there is a choice to use `Default` values or to enter values, the user can always select the choice to enter values and will be shown the defaults. The user can then alter any default values as desired. It is recommended that the user run initially with all default values, using the default noise simulation (Noise 2) for a subsurface dike consisting of three prisms.

Variable types The user can examine the provided .html files for quick determination of the variable types for inclusion in any data file.

Publish Generating your results using the MATLAB `publish` capability immediately provides a summary of all your results, with plots of the resulting gravity anomaly, diagnostics for the solution, and a plot of the point mass distribution.

Diagnostics The user can chose to illustrate the results just using a plot of the point mass distribution, the associated gravity anomaly and the computed values of the optimization functions with iteration number. If additional diagnostic analysis is required the user can chose to do linear regression analysis for the optimization functions and plot the results of the linear regression. There is also an option to export the results of the regression analysis to an excel spread sheet. The spread sheet summarizes the text of the Table generated in the Command window and included in the `publish` file.

Output All results are saved to a `.mat` file for later consideration, or plotting of results.

System Specific There may be some specific environment variables that will require modification. We point to the likely issues in the later discussion.

In the following discussion information on input parameters that need to be provided at each stage of the modeling are provided. The role of any script or function that does not require input parameters is noted in the summary provided in the Appendix A

2 GMD: Setting up the model and collecting data

Script `GMD.m` presents the user with a number of dialogue boxes to collect information on the model and generates a file that is used for the data inversion.

2.1 Input Parameters

The input parameters are described in Table 1. The requirements depend on whether the results are to be used for Real Data or for a simulated data set.

2.1.1 Real Data

It is assumed that the user has access to a measurement data set of the gravity, $\mathbf{d}_{\mathrm{obs}}$ and information about the survey volume that is saved in a datafile in a `.mat` format. `GMD.m` will present the user with a dialogue box requesting the name of the real data file, if the user indicates they are using real data. This file should contain

gse : Grid spacing in east direction. (scalar)

gsn : Grid spacing in north direction. (scalar)

nse : Number of stations in east direction. (scalar)

nsn : Number of stations in north direction. (scalar)

N : Total number of stations. (scalar)

Estat : East-coordinate of stations. (nse by 1 array)

Nstat : North-coordinate of stations (nse by 1 array)

Astat : Depth coordinate of stations((N by 1 array))

surveyvol : sizes in east, north and depth, with coordinate of the origin (1 by 6 cell array)

Sd : Standard deviation of the noise given for real data (m by 1 array)

dobs : Measured data for real data. (N by 1 array)

If any of the noted parameters are not found the user will be prompted via dialogue boxes to augment the information. For example if `surveyvol` is not given, the user is prompted for the values. The same applies for the noise distribution `Sd` The parameters of the survey volume for plotting are also saved. All data is saved to the file `AllRealData.mat` for input to `IGUG.m`.

Question	Options	Variable
Real Data	Yes (No)	
Name of File	Give Name of data file with real data	datafile
Survey Area Parameters	Yes (No)	
Size in East Direction (meters)	2000	East2=surveyvol{1}
Size in North Direction (meters)	1500	North2=surveyvol{2}
Size in DepthDirection (meters)	1000	Depth2=surveyvol{3}
East coordinate at the origin	0	East1=surveyvol{4}
North coordinate at the origin	0	North1=surveyvol{5}
Depth coordinate at the origin	0	Depth1=surveyvol{6}
Noise parameters	required if Sd not given for real data	
$(t1, t2)$	$t1\|(\mathbf{d}_{\mathrm{obs}})_i\| + t2\|\mathbf{d}_{\mathrm{obs}}\|$	Sd
Structure and Noise	Yes (No) for simulations	
Number of prisms	3	noc
Grid spacing in East Direction (meters)	50	gse
Grid spacing in North Direction (meters)	50	gsn
Pick a noise level	1 : 4 default 2	noise
Prism Parameters	Required for simulations: here 3 loops	(loops noc times)
First East Coordinate (meters)	$[700, 700, 700]$	East1C=prismdef{1}
Second East Coordinate (meters)	$[1300, 1300, 1300]$	East2C=prismdef{2}
First North Coordinate (meters)	$[800, 600, 400]$	North1C=prismdef{3}
Second North Coordinate (meters)	$[1200, 1000, 800]$	North2C=prismdef{4}
First DepthCoordinate (meters)	$[100, 250, 400]$	Depth1C=prismdef{5}
Second Depth Coordinate (meters)	$[250, 400, 550]$	Depth2C=prismdef{6}
Density of prism $grams/cm^3$	1	Density=prismdef{7}

Table 1: Input Parameters for Creating the Data Set using `GMD.m`. The resulting parameters are then used for `IGUG.m`

2.1.2 Simulated Data

In this case the user picks the parameters that are used for the simulations. This includes specifying the survey volume, grid spacing in east and north directions, the number of stations in east and north directions. The total number of stations is calculated. In the simulations it is assumed that all stations are on the surface, thus `Astat=0`. All data as described for the Real Data are generated. If any question is answered with "use default", default values will be used. For simulated data the user must provide coordinates for the subsurface structure. This can consist of `noc` blocks, for each of which the user provides the density and the defining sizes and coordinates. For Simulations the user has the choice of three given noise levels. Given the subsurface structure and the volume, the gravity response is calculated using `OnePrismResponse.p`. All data is saved to the file `DataNn.mat`, where n is the chosen noise level.

2.1.3 Plotting

For both real and simulated data `GMD.m` provides a plot of the obtained survey volume, showing the underlying subsurface structure determined by entering the information on the subsurface prisms. This plot can be used by the user to verify that appropriate inputs have been provided. The gravity anomaly $\mathbf{d}_{\mathrm{obs}}$ is also illustrated for both measured and simulated data. For measured data this assists the user in assessing that data has been correctly provided. For simulated data, the noise free gravity anomaly is also shown.

3 IGUG: Inversion of the data

Script `IGUG.m` contains a number of scripts/functions to determine the input data, the parameters to use for the genetic algorithm, find the solution of the inverse problem, and present options for the analysis of the results.

A data file generated from `GMD.m` is imported. The data flow includes the following scripts and functions.

1. `PlotDomainStructure.m` Function which plots the survey volume to verify that the data has been correctly imported. Will also show the subsurface structure for simulations.

2. `PlotObsData.m` Function to show the gravity anomaly. Also used to verify that data has been correctly imported.

3. `OptimParameters.m` This is the script that collects all the data used for the optimization.

4. `FindtheSolution.m` This is the script that finds the solution for the given configuration and parameters. It uses all the Private files and saves the results of the genetic algorithm.

5. `PredictedData.p` Function that generates the predicted anomaly for the given solution.

6. `Diagnostic_Results.m` Script to do regression analysis and plotting, dependent on user preference. Always shows comparative data for different choices of λ, when solved for multiple λ.

   `Regression_Analysis.m` Do linear regression analysis for the Θ, Φ, Γ curves as a function of the iteration number, using `datafit_linear.m`

   `PlotDomainStructure.m` Plot the survey volume and anomaly if exists.

   `ScatterPlotResults.m` Show the mass point distribution within the survey volume.

4 Script OptimParameters

Collects the parameters used to apply the genetic algorithm as noted in Table 2. It is possible to run this script without entering any parameters. Then the algorithm uses default values as summarized in Table 2.

Question	Default	Variable
Genetic Algorithm Parameters	Yes (No)	
Population Size	100	PopSize
Max Generations	200	GenNum
Cross Over Percentage	0.7	CP
Extra Range Factor for Crossover	0.4	Errf
Mutant Percentage	0.3	MP
Mutation Rate	0.1	μ
Selection Pressure	8	β
Mass points	20	M
Search Limits on Parameters	Yes (No)	
Minimum total mass	$70e9$	MinValm
Minimum in East Direction	400	MinValE
Minimum in North Direction	100	MinValN
Minimum in Depth Direction	20	MinValD
Maximum total mass	$150e9$	MaxValm
Maximum in East Direction	1600	MaxValE
Maximum in North Direction	1400	MaxValN
Maximum in Depth Direction	1000	MaxValD
Regularization parameter	Yes (No)	λ_i
Select multiple λ_i	0.2	$1 \leq i \leq 12$

Table 2: Input Parameters used for the Genetic Algorithm. All parameters can be selected as the defaults given. (No)

5 Script Diagnostic_Results

This script can be used to assist in interpretation of the results of the genetic algorithm. If all selections are "No", then the script will plot $(k, \log(\Gamma(k)))$, $(k, \Phi(k))$ and $(k, \log(\Theta(k))$ for each choice of λ. The resulting point mass distribution will be provided within the survey volume. Also a table of results are provided summarizing the final values of Γ, $\Phi/(N+\sqrt{2N})$ and Θ at the final iteration, as well as the selected iteration is noted and choice for λ. Finally, the user has the option as to whether to export the results of the table to a spreadsheet.

Question	Option	Impact
Regression Analysis	Yes (No)	Do linear regression analysis for Γ, Φ and $\log(\Theta)$
Select Range	Yes (No)	Select the range $[kstart, kend]$ for the regression
Save Figures	Yes (No)	Print the figures
Spread sheet	Yes (No)	Save results to a spreadsheet

Table 3: Diagnostic parameters for interpreting results. All parameters can be selected as "No".

6 Script Plot_Results

This script can be used to select results from a data file for plotting. It is recognized that after the user has studied the results obtained via Diagnostic_Results for all choices of λ, it may be convenient to do plots of single results for a chosen value of λ. Thus the user can run this script multiple times to generate a set of plots for a given λ, including the resulting gravity anomaly, the mass point distribution within the survey volume, and results of the regression analysis for Γ, Φ and Θ. Example titles are provided, but it is expected that the user will modify as appropriate.

Question	Option	Impact
Choice of λ	index i	selects λ_i for plotting
Regression Analysis	Yes (No)	Do linear regression analysis for Γ, Φ and $\log(\Theta)$
Mass Point Distribution	Yes (No)	Plot the mass point distribution and a slice view
Gravity Anomaly	Yes (No)	Plot the gravity anomaly

Table 4: Plotting of results for chosen λ. If all answers are "No" then this script does nothing.

Appendix A Directory Contents

High Level Directory containing

GMD.m Script to generate data.

IGUG.m Script to implement the genetic algorithm.

Open Directory Contained in High Level Directory and containing these files.

PlotDomainStructure.m Function. Plot the survey volume and for simulations also the subsurface structure.

PlotObsData.m Function. Plot the gravity anomaly.

OptimParameters.m Script. Set up the parameters to use for the genetic algorithm.

FindtheSolution.m Script. Implement the genetic algorithm for the given parameters and volume.

DiagnosticResults.m Script. Illustrate the results and do linear regression.

Regression_Analysis.m Function. Used to do the linear regression.

datafit_linear.m Function. Does linear data fit (regression) for a vector pair $(\mathbf{x}, \mathbf{y})$.

ScatterPlotResults.m Function Show the mass point distribution within the survey volume.

Plot_Results.m. Script. Use results file and plot chosen results.

Private Directory Contained in High Level Directory and containing these files.

OnePrismResponse.p Function that calculates gravity response for one structure.

Crossover.p Function. Apply crossover.

Equidistance.p Function. Calculate equidistance for the given population.

Fitness.p Function. Calculate the fitness of the given population.

InitialPopulation.p Function. Generate the initial population.

Mutate.p Function. Mutate the population.

PredictedData.p Generate predicted gravity anomaly for the current solution.

RWSelection.p Function. Roulette wheel selection of population.

Appendix B System Dependent Issues

File	Function	Issue	Fix
Diagnostic_Results	`writefile`	Datatype .xls	Chose a suitable file type.
Diagnostic_Results	`suptitle`	Interpretation of $\Theta(p)$, $\Gamma(p)$ and $\Phi(p)$.	Change the title
Any file doing plotting	`set(gcf)`	Figure placement	Change the positions.

Table 5: There are a number of functions that may need modification dependent on the local installation of MATLAB, the operating system, and MS Office Excel. These are noted here.

Appendix C Updates

February 21, 2019 Altered diagnostics to do exponential data fitting for Γ rather than linear data fitting.

References

[Bijani et al.(2015)] Bijani, R., Ponte-Neto, C. F., Carlos, D. U., Silva Dias, F. J. S., 2015. Three-dimensional gravity inversion using graph theory to delineate the skeleton of homogeneous sources, Geophysics, 80, G53-G66.

[Vatankhah et al. (2018)] Vatankhah, S., Ardestani, V. E., Niri, S. S., Renaut, R. A., Kabirzadeh, H., 2018. IGUG: A MATLAB program for 3-D inversion of gravity data using graph theory. Submitted