# Application of Compressive Sensing ideas to Static Electromagnetic Geophysics

EOSC 513: Course Project

THIBAUT ASTIC 90142150 thast@eos.ubc.ca

## Abstract

DC Resistivity (DCR) is a static electromagnetic surveying method. It is primarily used for near-surface characterization of buried infrastructure, mineral deposit, archeology and forensic investigation ([17]). The diagnostic physical property for DCR is electrical conductivity  $\sigma$  (or its inverse resistivity  $\rho$ ), which is mainly influenced by water and metal contents.

A set of strategies have been successfully used in SimPEG ([5]), an open source package for electromagnetic Geophysics, for inverting DCR data using a finite volume discretization. These strategies include the formulation of a L2-regularized objective function, which is then solved by updating a linearized version of the residual term thanks to an inexact Gauss-Newton step. The regularizer's weight is also regularly updated following a cooling strategy.

We propose in this project to apply some of the Compressive Sensing ideas to a DCR inverse problem.

Our first step will be to reformulate the inverse problem as a stochastic programming problem using simultaneous sources, as it has been applied recently in ([8]). We will present the impact of several strategies for the choices of the random matrix in term of its nature, size and updates.

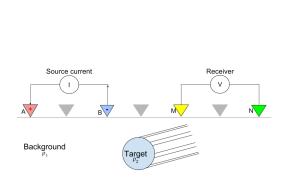
Our next step will be to promote sparsity in the Gauss-Newton subproblem in a transformed domain. For this purpose we will be using the SPG11 package presented in ([2]). This approach will help us to get additional highlights on our data such as for example the depth of investigation by exploring the possible models space in a different way.

Our open-source code has been made publicly available through Github, to ensure an easy reproducibility of the results presented in the final report as this has been a growing need in Geophysics and Sciences in general ([13]).

#### 1. Introduction

Characterizing the Earth's subsurface is one of the key questions in Earth Sciences, either to image the Earth's core, anticipating pollution migration, finding minerals resources etc. Geophysics goal is to image these structures remotely by measuring Earth's response to different fields.

One of the main group of methods used in Geophysics are the electromagnetic methods ([16]). With this project we are going to particularly focus on a static Electromagnetic method, DC Resistivity (DCR, figures 1 and 2). Electromagnetic phenomenons are governed by Maxwell's equations and hence are sensitive to contrasts in physical properties that can be interpreted in terms of geologic structure and fluid content. The diagnostic physical property for DCR is electrical conductivity  $\sigma$  (or its inverse resistivity  $\rho$ ). In a nutshell, DC Currents are injected into the ground thanks to a generator imposing a potential difference at the sources electrodes (electrodes A and B in figures 1 and 2). The path of the currents are determined by the electrical conductivity contrasts in the subsurface. Data are acquired either at the surface or in boreholes by measuring potential differences between measurements electrodes (electrodes M and N in figures 1 and 2).



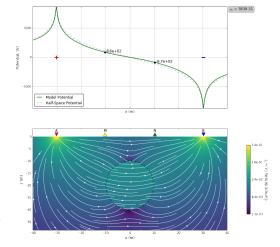


Figure 1: DCR system for a single source over a cylindrical target

Figure 2: DCR potential field at the surface and underground current flow

Maxwell's equation forms a complex system of Partial Differential Equations. In most cases, there is no analytic solution except for very simple problems (half-space, sphere in void). To forward model the response of complex structures, numerical techniques are required. We are using for this project the Python open-source package SimPEG ([5]) for the forward modeling of DCR data and perform L2-regularized inversions.

L2-regularized inversions is the most widely used type of inversions for diffusive electromagnetic and potential fields geophysics with a high rate of success ([7]). Recovered models using this type of inversion are smooth, which is compliant with the diffusive nature of the physics but not with the sharp geological contacts that are usually expected. Some progress have been made recently for potential fields inversions using L1-regularized inversions, through it needed implementations of Ad Hoc strategies ([6]).

Compressive Sensing is a popular topic in signal processing first introduced in ([3]). It aims at recovering sparse approximations of a signal in a transformed space from less measurements than required by the Shannon-Nyquist sampling theorem. its keys ideas are first to use a sparsifying transform to the model parameters and then work in this sparsifying space, second to generate coarse randomized sampling in the sparsifying domain, and at last to use a sparsity-promoting solver to recover a sparse approximations.

Despite being first thought for frequency measurements in ([3]), we propose in this project to apply some of these ideas to a static electromagnetic geophysics example. Random subsampling in DCR have been used in the past as for example in the DCR inversion with simultaneous sources case presented in ([8]) to reduce the number of right-hand sides. We will apply a similar stochastic approach to our case and present the impact of several strategies for the choices of the random matrix in term of its nature, size and updates. We will then promote sparsity in a transform space in the Gauss-Newton subproblems of the inversion both in the deterministic and stochastic formulation and compare the results. For this purpose we will be using the SPGl1 package presented in ([2]).

To ensure the reproducibility of this work, scripts to re-create most of the results generated for this project are available through Github at the address: https://github.com/thast/EOSC513

## 2. the DCR method

The formulation of the DCR problem and its discretization in finite volume have been done with extensive details in ([4]). We are reproducing here some of theirs visuals (figure 3).

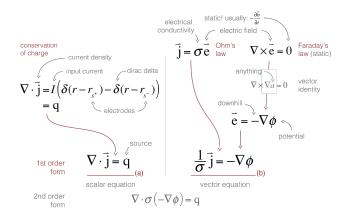


Figure 3: From Maxwell's equations to DC resistivity problem

Assuming Neumann boundaries conditions, we obtain the discretization of the DCR equation described in figure 3 in finite volume ([4]):

$$\operatorname{diag}(\mathbf{v})\mathbf{D}\mathbf{M}_f(\sigma^{-1})^{-1}\mathbf{D}^{\top}\operatorname{diag}(\mathbf{v})\boldsymbol{\phi} = \mathbf{q}$$
 (1)

$$A(m)\phi = q \tag{2}$$

With **D** the Divergence operator, **v** a vector containing the cells volume, and  $\mathbf{M}_f(\sigma^{-1})^{-1}$  a weighted averaging operator from cells center to faces (see [4] and [7]). the model m is usually chosen as the log of the electrical conductivity,  $m = ln(\sigma)$  to ensure the positivity of the recovered conductivity during the inversion process.

In a typical DCR survey, electrodes will be layout on a regular grid. Data will be acquired for several sequential sources. As a rule of thumb, the more spacing there is between the two sources electrodes, the deeper you investigate. Using several sources with different positions and spacing then give us information at several locations and depths. For several sources, the equation 2 can be written as:

for all sources 
$$q_i$$
:  $A(m)\phi_i = q_i$  (3)

or 
$$A(m)\Phi = Q$$
 (4)

With:

$$Q = \begin{bmatrix} & | & & & | \\ q_1 & q_2 & \dots & q_{N_s} \\ & | & & & | \end{bmatrix} \text{ and } \Phi = \begin{bmatrix} & | & & | & | \\ \phi_1 & \phi_2 & \dots & \phi_{N_s} \\ & | & & & | \end{bmatrix}$$
 (5)

and the observed data can be express as:

for all sources 
$$q_i$$
:  $d_i = PA(m)^{-1}q_i$  (6)

or 
$$D = PA(m)^{-1}Q \tag{7}$$

With P the projection matrix from the electric potential fields  $\phi$  to the receivers electrodes.

The matrix formulation presented in equations 4 and 7 will be useful at the conceptual level for formulating the simultaneous sources problem.

To learn more about the DCR method, We invite the reader to take a look at the website em.geosci developed at the University of British Columbia.

# 3. L2-Regularized Inversion

To reconstruct a model from observed data, we need to define an objective-function for our inversion problem. For DCR, the usual objective-function is:

$$f(m) = \frac{1}{2}||d(m) - d_{obs}||_2^2 + \frac{\beta}{2}||R(m - m_{ref})||_2^2$$
(8)

With  $d_{obs}$  a vector containing our observed data and d(m) a vector containing our modeled data with equation 6, the first term represents the data fitting objective. The second term is a Tikhonov-style regularizer ([15]) that contains a-priori information about the model, such as the reference model  $m_{ref}$ . The matrix R usually includes the Identity for smallness and first or higher order gradients for directional smoothness.  $\beta$  is the regularizer's weight. This term is iteratively cooled down during the inversion process, giving more and more weights to the data fitting term. Our goal is then to find a model m that minimize the value f. As this is an underdetermined nonlinear problem, an iterative scheme is use to solve this optimization problem. We look for an update  $\delta m$  such that we decrease enough the objective-function value:

$$f(m+\delta m) = \frac{1}{2}||r(m+\delta m)||_2^2 + \frac{\beta}{2}||W(m+\delta m - m_{ref})||_2^2$$
 (9)

s.t. 
$$f(m + \delta m) \le \gamma f(m)$$
 with  $\gamma \ge 1$  (10)

With the residual: 
$$r(m) = d(m) - d_{obs}$$
 (11)

To solve this problem, we first linearize the system:

$$f(m+\delta m) = \frac{1}{2}||r(m) + J(m)\delta m||_2^2 + \frac{\beta}{2}||W(m+\delta m - m_{ref})||_2^2$$
(12)

With J the Jacobian matrix of the data residual r ([7]).

We then obtain by derivation the Gauss-Newton system, whose solution gives us the Gauss-Newton update  $\delta m$ :

$$H\delta m = -g \tag{13}$$

With: 
$$H = J^T J + \beta W^T W$$
 (14)

and: 
$$q = J^T r(m) + \beta W^T W(m - m_{ref})$$
 (15)

We finally update our previous model  $m_k$  with an additional line search

$$m_{k+1} = m_k + \alpha \delta m \text{ s.t. } f(m_{k+1}) < \gamma f(m_k) \text{ with } \gamma \ge 1$$
 (16)

We applied this scheme to an toy example problem (figure 4). This example is composed of 23 dipoles sources of various length. Every sources share the same dipoles receivers for a total of 276 measurements. The recovered model is presented in figure 5. The overall position and polarity of the anomalies is well recovered but, as expected, the result is quite smooth and it is hard to get a precise idea of the initial shapes.

We will then try to use some ideas from Compressive Sensing to our case to sparsify the recovered model in a transform space that might improve the anomalies locations and values in the recovered models.

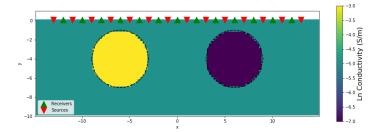


Figure 4: Initial model and survey

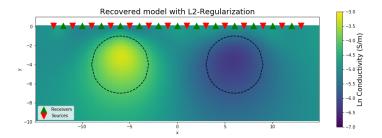


Figure 5: Recovered model with L2 Regularization

# 4. Basics ideas of Compressive Sensing

The first goal of Compressive Sensing is to solve the underdetermined linear problem:

$$\min_{x} ||x||_0 \text{ s.t. } Ax = b \tag{17}$$

Compressive Sensing theory is telling us that if A is an incoherent matrix and x is sparse, x can be recovered exactly under certain conditions. If the original x is not sparse but can be represent sparsely in some transform space through a linear transform S, we can rewrite the original problem 17 as:

$$\min_{x} ||x||_0 \text{ s.t. } RMS^H x = b$$
 (18)

with S the sparsifying transform, M the measurement matrix and R the restriction matrix. According to Compressive Sensing theory, We need the rows of M and S to be incoherent, which means their inner-products are small. However this is still an NP-hard problem. [3] showed that, under certain conditions, the L1 minimization problem (equation 19) can still recover a robust approximation of x, with less measurements than required by the Shannon-Nyquist sampling theorem by using coarse randomized sampling in the sparsifying domain.

$$\min_{x} ||x||_1 \text{ s.t. } RMS^H x = b \tag{19}$$

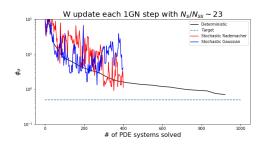
In the following project, we have used the SPGl1 package ([2]) to solve the problem presented in equation 19

## 5. Simultaneous Sources

We first started by investigating random simultaneous sources in the case of DCR survey. We can formulate the inverse problem shown in equation 8 as a stochastic problem, following the work done in ([8]), by randomly mixing the original sources into a smaller number of simultaneous sources. Reducing the number of sources can make computing the Gauss-Newton update much computationally efficient as for each step we need to solve  $2N_s$  Partial Differential Equations systems, with  $N_s$  the number of sources. We generate the random simultaneous sources following:

$$f(m) = \frac{1}{2} ||PA(m)^{-1}QW - DW||_F^2 + \beta R(m)$$
 (20)

With W a random Gaussian or Rademacher matrix of size  $N_s \times N_{ss}$  with  $N_{ss} << N_s$ ,  $N_{ss}$  being the number of final recombined sources and  $N_s$  the original number of sources. The Gauss-Newton step subproblem's size is then reduced by a factor  $\frac{N_s}{N_{ss}}$ 



**Figure 6:** Reduction to 1 source, update W each GN step

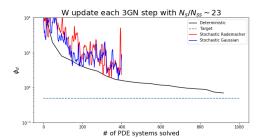


Figure 8: Reduction to 1 source, update W each 3 GN step



Figure 7: Reduction to 5 sources, update W each GN step



Figure 9: Reduction to 3 sources, update W each 3 GN step

Without any update to W during the inversion process, we have seen that we converge to a local minima but that the recovered model, when compared to the full data set, still present a very high misfit. We have then experiment several strategies for updating W (figure 6, 7, 8, 9) and their effects on the convergence rate for the data misfit. The parameter  $\beta$  was still cooled down to mimic the strategy used during the deterministic inversion. The different graphs represent the data misfit of the model recovered after each Gauss-Newton step compared to the whole data set, not the simulated data obtained through the simultaneous sources. We represent in blue the tests with Gaussian matrix, in red the tests with Rademacher matrix and in black the deterministic case. The dashed line indicates the misfit target based on the noise level of the data. We observe that taking a Gauss-Newton step on a particular realization of a simultaneous sources does not guarantee that this will decrease the misfit with the overall dataset. However we can see that the stochastic curves

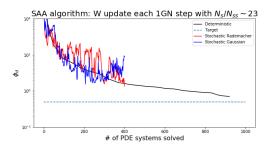


Figure 10: SA algorithm: Keep in memory the previous update

roughly follow the trend of the deterministic case. As expected, the convergence curves obtained when we less reduce the number of sources (going from 23 original sources to 5 instead of just 1) are smoother and closer to the deterministic case.

We also took a look at a stochastic algorithm that keep in memory the previous updates, as presented in ([8]) (figure 10), named Stochastic Approximation (SA). We can observe that this improves slightly the behavior of the convergence curves compared to figure 6 but the same flows are still present. It is very likely that our test example is too small to benefit from such an approach. It is also possible that, as we are in a static case and thus dealing with potential fields, that the random mixing of these fields create unexpected patterns, emphasizing much more the sensitivity of certain areas over others that get masked.

Overall we would only recommend using this approach for huge dataset for which computing the Gauss-Newton update based on the full original dataset would not be possible memory-wise and with better chance to have redundant data.

# 6. Promoting Sparse Gauss-Newton update

In this section, we are aiming at solving system similar to equation 19 defined by Compressive Sensing using the SPGl1 solver [2].

#### 6.1 Formulating the subproblem

We can rewrite our original linearized problem from equation 12 in a similar form than in equation 19 without the L2 penalty but a L1 constrain in some sparsifying domain:

$$\min ||\delta x||_1 \text{ s.t } ||J(m)S^H \delta x + r(m)||_2^2 < \tau$$
 (21)

With

$$\delta x = S\delta m \tag{22}$$

Note that if we still have additional information such as a good reference model or weights, we can still add it as an L2 term in the above formulation. Bounds on the physical properties are also possible, through it would present its own challenge to implement it successfully.

#### 6.2 Choosing a sparsifying transform

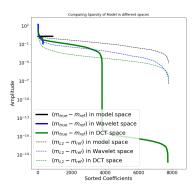


Figure 11: Sparsity of the true model (plain) and recovered model (dashed) (minus reference model) in different spaces

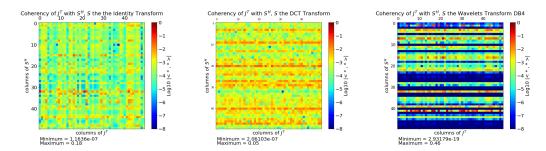


Figure 12: Coherency  $J^T$  with Identity

Figure 13: Coherency  $J^T$  with DCT

Figure 14: Coherency  $J^T$  with the WVT4

Based on classic Compressive Sensing, choosing the sparsifying transform requires two conditions. First the model we are trying to recover can be decompose or at least approximate by a sparse representation in the chosen space (figure 11). Second, to ensure some guarantee on the quality of the recovery, the columns of  $J^T$  and  $S^H$  in equation 21 should be incoherent.

We choose here to test the Identity transform (no transform), because this is in the end the domain of interest. We also tested the Direct Cosine Transform (referred later as DCT), as the DCR problem is a weighted Laplacian and that the eigenfunctions of a Laplacian are cosine and sine functions. Intuitively it then seems that just a few coefficients in this space should be able to explain the data. Finally we tested the 4th order Daubechies wavelet transform (WVT4) as a compromise between localized but still smooth model update. We also tested the lower orders wavelet transforms but we got our best results with the 4th order ones.

Based on figure 11,from the true model point of view, both Identity and Wavelet seem like potential candidates for S while the DCT do not help much sparsifying the model. From the DCR Jacobian point of view, the DCT presents the most constant and lowest maximum of coherency (figure 13). This was expected, for the reasons mentioned above. Interestingly, the WVT4 present a lot of very small coherency coefficients with few very high coefficients (figure 14). The Identity transform presents an almost uniform coherency that is fairly higher than the DCT but lower than the maximum for the WVT4 (figure 12).

#### 6.3 Recovered models

We compare here the models obtained by using different sparsifying transform such as discussed in the previous subsection.

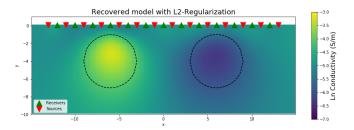


Figure 15: Recovered model with L2 regularization

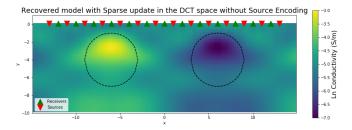


Figure 16: Recovered model with DCT as S

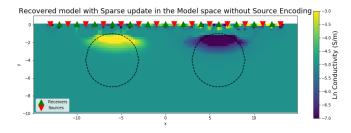


Figure 17: Recovered model with Identity as S

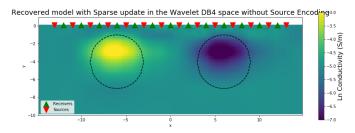


Figure 18: Recovered model with WVT4 as S

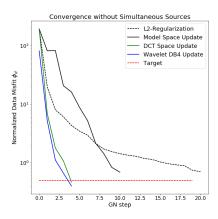
Interestingly, each model are quite different and reach (or close to reach) the target data misfit (figure 19).

For the sparse GN update in the model space (Identity transform), SPG11 got more difficulties to solve the problem (figure 19) and each iteration took a long time to compute compared to update in other spaces. The recovered model (figure 17) is quite compact but noisy. When looking at the updates, even without simultaneous sources, we realize these updates are not that sparse (figure 20).

For the sparse GN updates in the DCT space (DCT transform), SPG11 reach the target misfit in just few GN steps (figure 19), which we can relate to the small coherency between of J and S. The recovered model still displays the periodicity of the basis function, which has nothing to do with the true model (figure 16).

Finally, despite a few high coherency coefficients between J and S, the Wavelet transform with 4th order Daubechies wavelets gave us a satisfying result, both in term of the number of iterations to reach the target data misfit (figure 19) and in term of the recovered model (figure 18). The recovered model is still smooth but anomalies are must less spread out compared to the L2-regularized model. The recovered physical properties are also much closer to those of the true model.

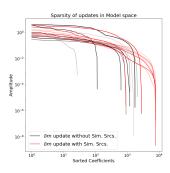
We do not have a definite answer to why the wavelet transform works that well despite some high-coherency coefficients. A hypothesis to test would be to evaluate if our Jacobian J and the wavelet transform are "asymptotically incoherent" instead of fully incoherent. Under this conditions, it has been shown recently that most of the Compressive Sensing results still hold ([1]).

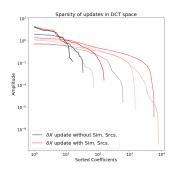


**Figure 19:** Convergence of the data misfit for each type of Gauss-Newton update

#### 6.4 Sparse Gauss-Newton update with Simultaneous Sources

In this section, we present to the audience an unexpected problem we encountered while combining the sparse Gauss-Newton update approach with the simultaneous sources. Our expectation would have been that, as when we are decreasing the number of sources we are also decreasing the available information at each iteration, this would also simplifies sparsifying the update. What happened is the exact opposite. By mixing the sources together into a smaller number of final sources, SPGl1 became much slower to solve each subproblem. Worse, the updates where no longer sparse anymore, or not as sparse as before, whatever transformed we used (figure 20, 21, 22). For now, one possible explanation refers back to the discussion about the simultaneous sources results, that our example might be too small to use a stochastic approach or unevenly highlight certain areas over others.





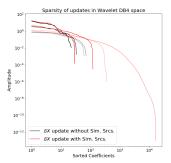


Figure 20: Sparsity of updates Figure 21: Sparsity of update Figure 22: Sparsity of updates with Identity as S with DCT as S with WVT4 as S

#### 7. Discussion and Conclusion

The tests we performed on our toy example for simultaneous sources have shown a similar or worse rate of convergence, in term of number of PDE systems solved, compared to using the full system in a deterministic process. However each step is a lot less memory-expansive, making it useful if yours systems are too big to be efficiently stored. It is possible that we may have obtained better results on a larger example with more redundancy of information in the acquired data. When used in combination with the sparse update approach, simultaneous sources have make things very difficult for SPG11 to solve, in reverse to our initial intuition that decreasing the level of information would make the update easier and sparser.

At first, it seemed we had an antagonization for choosing a sparsifying transform that both allows a sparse representation of the expected model and that is incoherent with the Jacobian of the DCR problem. The Direct Cosine Transform was appropriate with the diffusive nature of the physics but did not allow a good reconstruction of the model. An identity transform was not enough incoherent to be efficiently solved and to ensure the quality of the reconstruction. Finally, a 4th order Daubechies wavelet transform seemed to be able to bridge the gap between the model and the physics by allowing smooth but very localized model updates despite some very high coherent coefficients with the DCR Jacobian. Our hypothesis to explain why the wavelet transform is working is that it may be asymptotically incoherent with the DCR Jacobian, which would still be a sufficient condition to preserve the properties of Compressive Sensing.

We did not study the influence of the  $\tau$  parameter for the data fitting term in SPGL1 or the strategy to set it. The few tests we performed did not seem to change much the final recovered model but only the number of Gauss-Newton updates to reach the desired target data misfit and the time necessary to compute each of the update. The smaller  $\tau$  was, the fewer Gauss-Newton steps was needed but each step was more expansive to compute. There could be then an ideal value that optimize the time taken by the inversion but it could be highly problem-specific.

Finally, to ensure the reproducibility of this work, scripts to re-create most of the results generated for this project are available through Github at the address: https://github.com/thast/EOSC513

# 8. Acknowledgments

We would like to thanks Seogi Kang and Devin Cowan, all students at UBC - Geophysical Inversion Facility for sharing theirs experiences during the making of this project.

## References

- [1] Adcock, Ben, et al. (2017) "Breaking the coherence barrier: A new theory for compressed Sensing." Forum of Mathematics, Sigma. Vol. 5. Cambridge University Press, 2017.
- [2] Berg E. van den and Friedlander M. P. (2008) "Probing the Pareto frontier for basis pursuit solutions", SIAM J. on Scientific Computing, 31(2):890-912, November 2008
- [3] Candes E., Romberg J., Tao T. (2006) "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information" IEEE Trans. Inform. Theory, 52(2): 489-509
- [4] Rowan Cockett, Lindsey J. Heagy, and Douglas W. Oldenburg (2016) Pixels and their neighbors: Finite volume. The Leading Edge, 35(8), 703706. doi: 10.1190/tle35080703.1 http://tutorials.simpeg.xyz/content/pixelsandtheirneighbors.html
- [5] Cockett, Rowan, Seogi Kang, Lindsey J. Heagy, Adam Pidlisecky, and Douglas W. Oldenburg (2015) "SimPEG: An Open Source Framework for Simulation and Gradient Based Parameter Estimation in Geophysical Applications." Computers and Geosciences, September 2015. doi:10.1016/j.cageo.2015.09.015.
- [6] Fournier D. (2015) A cooperative magnetic inversion method with Lp-norm regularization, M.Sc. Thesis
- [7] Haber E.(2014). "Computational Methods in Geophysical Electrodynamics"
- [8] E. Haber, M. Chung, and F. Herrmann (2012) An Effective Method for Parameter Estimation with PDE Constraints with Multiple Right-Hand Sides, SIAM Journal on Optimization, vol. 22, no. 3, pp. 739757, jul 2012.
- [9] Heagy J. lindsey, Rowan Cockett, Seogi Kang, Gudni K. Rosenkjaer, Douglas W. Oldenburg (2016) "A framework for simulation and inversion in electromagnetics" *Geophysics*
- [10] Luisier M., O. Schenk et.al. (2013) Fast Methods for Computing Selected Elements of the Green's Function in Massively Parallel Nanoelectronic Device Simulations Euro-Par 2013, LNCS 8097, F. Wolf, B. Mohr, and D. an Ney (Eds.), Springer-Verlag Berlin Heidelberg, pp. 533544, 2013.
- [11] Schenk O., M. Bollhfer, and R. Rmer, On large-scale diagonalization techniques for the Anderson model of localization. Featured SIGEST paper in the SIAM Review selected "on the basis of its exceptional interest to the entire SIAM community". SIAM Review 50 (2008), pp. 91-112.
- [12] Schenk O., A. Wchter, and M. Hagemann, Matching-based Preprocessing Algorithms to the Solution of Saddle-Point Problems in Large-Scale Nonconvex Interior-Point Optimization. Journal of Computational Optimization and Applications, pp. 321-341, Volume 36, Numbers 2-3 / April, 2007.
- [13] Society of Exploration Geophysicists Reproducible research: Geophysics papers of the future 2016 http://library.seg.org/page/Geophysics/reproducible-research:-geophysics-papers-future?mobileUi=0
- [14] Telford, W.M., Geldart, L.P. and Sheriff, R.E. (1990) Applied Geophysics, 2nd Edition: Cambridge University Press, 1990.

- [15] Tikhonov, A.N., Arsenin, .V.Y. (1977) "Solution of Ill-posed problems", Winston and Sons, Washington, 1977
- [16] S. H. Ward, C. W. Hohmann (1988). "Electromagnetic Theory for Geophysical Applications", in Nabighian, M. N., Ed., Electromagnetic Methods in Applied Geophysics: Society of Exploration Geophysics.
- [17] Yuval, and D. W. Oldenburg (1996) DC resistivity and IP methods in acid mine drainage problems: results from the Copper Cliff mine tailings impoundments Journal of Applied Geophysics, 34, no. 3, 187-198, 1996