Iteratively Regularized Gauss Newton With Variable Weighting

Rosemary Renaut
Taufiquar Khan and Alexandra Smirnova

Arizona State University, Clemson, and Georgia State University

Copper Mountain April 2006

Inverse Problem in Optical Tomography

$$-\nabla \cdot D(\mathbf{x})\nabla u(\mathbf{x};\mathbf{q}) + \mu_a(\mathbf{x})u(\mathbf{x};\mathbf{q}) = s(\mathbf{x}), \tag{1}$$

- **Goal** Estimate the model parameters $\mathbf{q} = (D, \mu_a)^T$,
  - Coefficient of diffusion D
  - Coefficient of absorption μ_a
- **Provided:** Boundary Measurements of
  - $u(\mathbf{x};\mathbf{q}) = \mathbf{g}_i, i = 1 \dots n_m$ for
- **Given:** input signals
  - $s(\mathbf{x}) = (s_j), j = 1 \dots n_s$
- **References** The forward model (1) is described elsewhere, [5, 2] and others.

Nonlinear Least Squares Optimization

- **Minimize** residuals over all sources and measurements.

$$F_{ij} = C_{ij} - g_{ij}$$

- **Fidelity Term**

$$\begin{aligned} \Psi(\mathbf{q}) &= \frac{1}{2} \arg\min_{\mathbf{q}} \|F\|_F^2, \\ &= \frac{1}{2} \arg\min_{\mathbf{q}} \sum_{j=1}^{n_m} \sum_{i=1}^{n_s} (C_{ij}(u(\mathbf{q})) - g_{ij})^2. \end{aligned} \quad (2)$$

- **Nonlinearity** C_{ij} obtained by solving (1) - the forward problem
- **Ill-Posed** Dimension of $\mathbf{q}$, $2n > n_s m_s$, where n determines discretization in space. (Grid dimension)

Quasi-Newton

- **Standard Well-Known Approach** Based on Newton step

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}. \tag{3}$$

scalar $\alpha^{(k)} \leq 1$ is line search parameter in search direction $\mathbf{p}^{(k)}$

$$(\nabla^2\Psi(\mathbf{q}^{(k)}))\mathbf{p}^{(k)} = -\nabla\Psi(\mathbf{q}^{(k)}), \tag{4}$$

- Non-linear Least Squares uses approximation

$$\nabla^2\Psi(\mathbf{q}^{(k)}) = K^*(\mathbf{q}^{(k)})K(\mathbf{q}^{(k)}) + A(\mathbf{q}^{(k)}) \approx K^*(\mathbf{q}^{(k)})K(\mathbf{q}^{(k)}), \tag{5}$$

$K(\mathbf{q}) = F'(\mathbf{q})$ and $\nabla\Psi(\mathbf{q}) = K^*(\mathbf{q})F(\mathbf{q})$.

- But $K^*(\mathbf{q})K(\mathbf{q})$ can not be assumed to be continuously invertible.

Regularization

- Levenberg-Marquardt Method - regularization parameter η

$$(K^*(\mathbf{q}^{(k)})K(\mathbf{q}^{(k)}) + \eta^{(k)} I)\mathbf{p}^{(k)} = -K^*(\mathbf{q}^{(k)})F(\mathbf{q}^{(k)}) \qquad (6)$$

$$\eta^{(k)} = \min\{1, \|F(\mathbf{q}^{(k)})\|\}$$

- Solve for example without forming $K^*(\mathbf{q}^{(k)})K(\mathbf{q}^{(k)})$ minimize

$$\| \begin{pmatrix} K(\mathbf{q}^{(k)}) \\ \sqrt{\eta^{(k)}} \end{pmatrix} \mathbf{p}^{(k)} - \begin{pmatrix} -F(\mathbf{q}^{(k)}) \\ 0 \end{pmatrix} \|_2^2, \qquad (7)$$

- **Goal: Effectively** approximate the pseudoinverse

$$(K^*K + \eta I)^{-1} K^* y \rightarrow K^\dagger y \,\mathrm{as}\, \eta \rightarrow 0 \,\forall y \in \mathcal{R}(K) \qquad (8)$$

Iteratively Regularized Gauss Newton [1]

- In (6) introduce prior information $\mathbf{q}_r$ and solve

$$(K^*K + \eta^{(k)}I)\mathbf{p}^{(k)} = -(K^*F + \eta^{(k)}(\mathbf{q}^{(k)} - \mathbf{q}_r)) \qquad (9)$$

- Equivalently change cost functional

$$\Psi = \frac{1}{2}(\arg\min_{\mathbf{q}} \|F\|_F^2 + \tau R(\mathbf{q})), \; R(\mathbf{q}^{(k)}) = \|(\mathbf{q}^{(k)} - \mathbf{q}_r)\|_2^2, \qquad (10)$$

- Yields LS update and absorbing τ in $\eta^{(k)}$

$$\| \begin{pmatrix} K(\mathbf{q}^{(k)}) \\ \sqrt{\eta^{(k)}} \end{pmatrix} \mathbf{p}^{(k)} + \begin{pmatrix} F(\mathbf{q}^{(k)}) \\ \sqrt{\eta^{(k)}}(\mathbf{q}^{(k)} - \mathbf{q}_r) \end{pmatrix} \|_2^2, \qquad (11)$$

Local Convergence Results [3]

- Let $\mathbf{q}^*$ be a minimizer of the fidelity term (2) then

$$\|\mathbf{q}^{(k)} - \mathbf{q}^*\| = O((\eta^{(k)})^\nu), \quad \frac{1}{2} \leq \nu < 1 \tag{12}$$

  assuming
  - F' is Lipschitz Continuous
  - Smoothness : $\mathbf{q}^* - \mathbf{q}_r = (K^*(\mathbf{q}^*)K(\mathbf{q}^*))^\nu v$ where $v \in \mathcal{N}(K(\mathbf{q}^*))^\perp \setminus \{0\}$ is small
  - $$\eta^{(k)} > 0, \quad 1 \leq \frac{\eta^{(k)}}{\eta^{(k+1)}} \leq t, \quad \lim_{k\to\infty} \eta^{(k)} = 0. \tag{13}$$

- Under additional conditions (12) holds for $\nu < \frac{1}{2}$

Additional Conditions

- More general regularization

$$R(\mathbf{q}^{(k)}) = \|L(\mathbf{q}^{(k)} - \mathbf{q}_r)\|_2^2, \tag{14}$$

- Yields LS update

$$\| \begin{pmatrix} K(\mathbf{q}^{(k)}) \\ \sqrt{\eta^{(k)}}L \end{pmatrix} \mathbf{p}^{(k)} + \begin{pmatrix} F(\mathbf{q}^{(k)}) \\ \sqrt{\eta^{(k)}}L(\mathbf{q}^{(k)} - \mathbf{q}_r) \end{pmatrix} \|_2^2, \tag{15}$$

- Permits imposition of smoothness conditions on $\mathbf{q}$ through L.
- Total Variation also fits this model when

$$\mathrm{TV}_\beta(f) = \int_{-\infty}^{\infty} \sqrt{f'(t)^2 + \beta}dt, \tag{16}$$

although $L = L(\mathbf{q})$.

ARIZONA STATE UNIVERSITY

Initial Local Convergence Results

- Let $\mathbf{q}^*$ be a minimizer of the fidelity term (2) then

$$\|\mathbf{q}^{(k)} - \mathbf{q}^*\| = O((\sqrt{\eta^{(k)}})) \tag{17}$$

assuming, in addition

- $L^T L(\mathbf{q}^* - \mathbf{q}_r) = \delta^2$ where δ is a given bound.
- $\mathcal{N}(K^*K) \cap \mathcal{N}(L^T L) = \varnothing$
- **Problems**
  - Methods still require choice of iterating sequence.
  - Conditions on proofs are limiting, require good guesses etc.
  - Assume **exact solves** for search direction!

Practicalities: Parameter Decompostion

- The specific Optical Tomography (1) contains distinct parameters, D and μ, of different order of magnitude and perhaps different properties.

$$\Psi = \frac{1}{2}(\arg\min_{\mathbf{q}} \|F\|_F^2 + \eta_1 R_1(\mathbf{q}_1) + \eta_2 R_2(\mathbf{q}_2)), \tag{18}$$

- Yields LS

$$\| \begin{pmatrix} K(\mathbf{q}^{(k)}) & \\ \sqrt{\eta_1} L_1 & 0 \\ 0 & \sqrt{\eta_2} L_2 \end{pmatrix} \mathbf{p}^{(k)} + \begin{pmatrix} F(\mathbf{q}^{(k)}) \\ \sqrt{\eta_1} L_1(\mathbf{q}_1 - \mathbf{q}_{1r}) \\ \sqrt{\eta_2} L_2(\mathbf{q}_2 - \mathbf{q}_{2r}) \end{pmatrix} \|_2^2. \tag{19}$$

1D Forward Simulations

Model Solve on interval $x \in [0, 43]$, where x is physical space.

Parameters Background values of physical parameters are $\mu = .006$ and $D = .55$ with discontinuity on $[1, 10]$, values $\mu = .012$, $D = .275$

Discretization Use a fine grid, eg $N_x = 153$, *avoid inverse crimes!*.

Sources 10 placed near the boundary, equally spaced between 0.1 and 1.5 and between 41 and 42.5.
$f(x) = 1/\sqrt{.02\pi}\exp(-40|x - x_s|^2)$

Measurements 2 per source one at each boundary.

Inverse Problem

Parameters Reference solution: $\mu = .006$ and $D = .55$. Data are initialized also at the reference.

Discretization Coarse grid, eg $N_D = 20, N_\mu = 40$

Parameter Space Finite element solver uses spline coefficients, $\mathbf{q}^T = (\xi, \varsigma)$ and not the physical values.

$$D^{N_D}(x) = \sum_{j=1}^{N_D} \xi_j b_j(x) \quad \mu^{N_\mu}(x) = \sum_{j=1}^{N_\mu} \varsigma_j b_j(x),$$

Matrices B_1 and B_2 $D(x_i) = d_i = (B_1\xi)_i$ and similarly for μ.

Regularization

- $\eta^{(k)} = \eta^{(0)}/k$, where $\eta^{(0)} = .0025$ typically.
- Use basis matrices to map spline coefficients to physical space, $L_i = B_i$.
- Notice that B_i are invertible.
- To account for difference in scale introduce weight on μ, $L_2 = wB_2$, obtained by observing ratio of cancer parameters is about $w = 50$. Then equal weight is given to both regularization terms.

General Convergence Conditions

- **Outer Iteration** terminated $\|F'(\mathbf{q}^{(k)})\| < TOL$ where TOL is a specified tolerance, typically 10^{-4}.
- **Search direction** is obtained using lsqr. (*Inexact solve prevents convergence*)
- **Line Search** use weak Wolfe Conditions

$$\Psi(\mathbf{q}^{(k)} + \alpha \mathbf{p}^{(k)}) \leq \Psi(\mathbf{q}^{(k)}) + \sigma_1 \alpha \nabla \Psi(\mathbf{q}^{(k)})^T \mathbf{p}^{(k)} \quad (20)$$

$$\nabla \Psi(\mathbf{q}^{(k)} + \alpha \mathbf{p}^{(k)})^T \mathbf{p}^{(k)} \geq \sigma_2 \nabla \Psi(\mathbf{q}^{(k)})^T \mathbf{p}^{(k)} \quad (21)$$

$$0 < \sigma_1 < \sigma_2 < 1 \quad (22)$$

- Standard: $\sigma_1 = .0001, \sigma_2 = .99$.
- When line search fails use steepest descent.
- Project negative values back to background.

Results: Discussion

- Convergence is very sensitive to choices of
  - Weight w : Earlier results with same codes could not converge without the weighting, for both D and μ.
  - Choice of $\eta^{(0)}$. It should neither be too small or too large. It is independent of the chosen grids.
  - Position of discontinuity:, near boundary (this is expected).
  - Using steepest descent if line search is unsuccessful after say, 5 inner steps. This reduces the total computation.
  - Finer grid for μ than for D.

Results: Graph

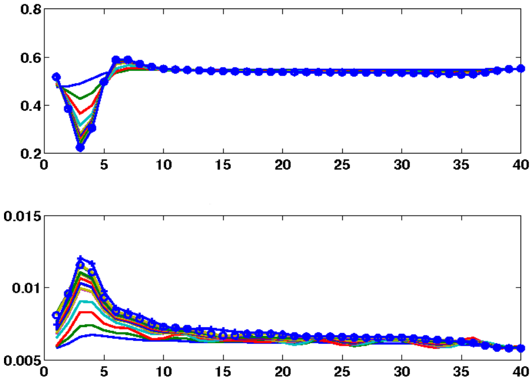

Figure: D above and μ below, interpolated to 40 points from $N_D = 20$, $N_\mu = 40$. Used $w = 25$, $\eta^{(0)} = .01$. 29 search directions and not converged

Results: Convergence Characteristics

Step	Ddiff	μdiff	grad(μ)	grad(D)	η
1	8.5e-02	2.7e-01	2.3e-01	1.7e-03	2.5e-03
2	1.5e-02	1.1e-01	1.1e-02	3.4e-04	1.3e-03
3	1.7e-02	1.1e-01	7.2e-03	1.6e-04	8.3e-04
4	8.3e-02	2.1e-01	2.5e-03	9.2e-05	6.3e-04
5	2.8e-02	2.0e-01	5.7e-02	7.9e-04	5.0e-04
6	2.6e-02	2.0e-01	9.0e-03	1.5e-04	4.2e-04
7	1.0e-02	7.6e-02	2.6e-03	5.1e-05	3.6e-04

Modification

- Use a decomposition approach
- Monitor ∇D and $\nabla \mu$ separately
- Iterate only on μ coefficients if D is converged to tolerance given. Fix D.
- Convergence results still apply, subproblem involves only μ.
- Restart with new $\eta^{(0)}$
- Saves significant computation - recall each search direction solves entire forward problem.

Results: Decomposition

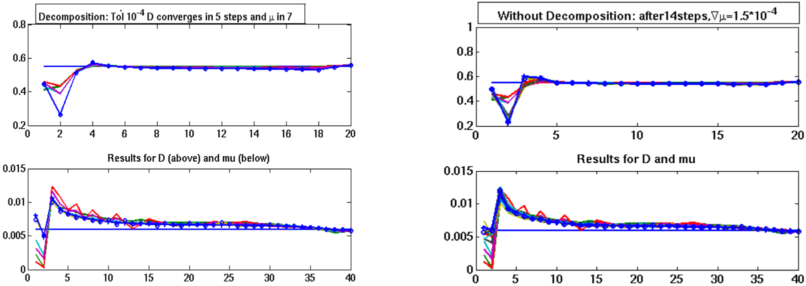

Figure: Left with decomposition, right without, tol=.0001, reached in total 7 steps on left, D converged in 5. On right after 14 steps and 24 inner line search steps, $\nabla\mu = .00015$.

Noise: 1%

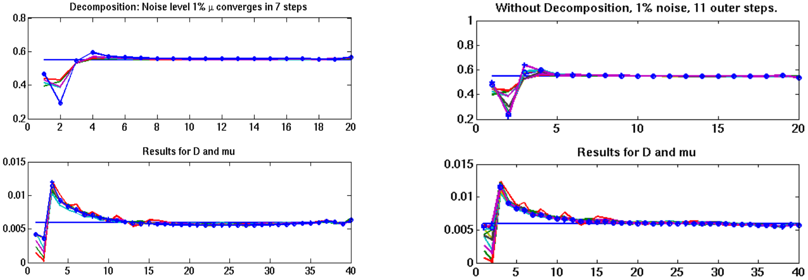

Figure: Left with decomposition, right without, tol=.0001, reached in total 7 steps on left, D converged in 5, on right after 11 steps. Neither converge with noise 2.5%

Observations and Future Work

- IRGN extend for more general regularizing conditions.
  - *Is there other work on this?*
- Implement TV and TK regularization with different L.
- Extend to 2D problem.
- Should be applied to some other interesting problems
- Current formulation uses specific LS solve:
  - Extend to use inexact inner solve - BFGS/Krylov etc
  - Convergence theory for inexact solves.
- Investigate robustness to noise levels in data[3, 4]
- Alternative IRGN use stabilizing iteration $q^{(k+1)} = q^{(0)} - \eta^{(k)}\mathbf{p}^{(k)}$
- Investigate improving D after decomposition - notice decomposition results are slightly inferior.

Bakushinsky, A. B., 1992, The Problem of the convergence of the iteratively regularized Gauss-Newton method, *Comput. Math. Meth. Phys.*, **32**, 1353-1359

Bakushinsky, A. B., Khan, T., and Smirnova, A. 2005, Inverse Problem in Optical Tomography and its Numerical Investigation by Iteratively Regularized Methods, *Journal of Inverse and Ill-posed Problems*, **13**,4, 1-14.

Blaschke, B., Neubauer, A. and Scherzer, O, 1997, On Convergence Rates for the Iteratively Regularized Gauss-Newton Method, *IMA Journal of Numerical Analysis*, **17**, 421-436.

Burger, M., and Kaltenbacher, B., 2006, Regularizing Newton–Kaczmarz Methods for Nonlinear Ill-Posed Problems, *SIAM J. Num Anal.*, **44**, 1, 153-182.

Khan, T. and Smirnova, A.,2005, 1D inverse problem in diffusion based optical tomography using iteratively regularized Gauss-Newton algorithm, *Applied Mathematics and Computation*, Elsevier Science Inc., **161**, 149-170.

Nash, S.G., and Nocedal, J., 1991, A Numerical Study of the Limited Memory BFGS methods and the truncated-Newton method for large scale optimization, *SIAM J. Optim.*, **1**, 358-372.

Zhu, C. & Byrd, R. & Lu, P. & Nocedal, J., 1997. Software for Large-scale Bound- constrained or Unconstrained Optimization, http://www-fp.mcs.anl.gov/otc/Tools/LBFGS-B/.