Approaching Linear Ill-posed Problems by an Entropic Projection Method

Elena Resmerita

Institute of Mathematics
Alpen-Adria Universität
Klagenfurt, Austria

Inverse problems: theory and statistical inference
October 28-29, 2016, Heidelberg

Introduction

Consider an ill-posed linear operator equation

$$Au = y$$

with $A : L^1(\Omega) \to Y$ bounded and Ω an open bounded domain in $\mathbb{R}^d$.

Aim: Recovering a nonnegative solution of the equation, when it exists.

Joint work with Martin Burger (University of Münster)

Entropy functionals

The (negative of the) Boltzmann-Shannon entropy $f : L^1(\Omega) \to (-\infty, +\infty]$ is defined as

$$f(u) = \begin{cases} \int_\Omega u(t) \log u(t)\, dt & \text{if } u \geq 0 \text{ a.e. and } u \log u \in L^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

The Kullback-Leibler functional $d : dom\, f \times dom\, f \to [0, +\infty]$ is

$$d(v, u) = f(v) - f(u) - f'(u, v - u),$$

$$d(v, u) = \int_\Omega \left[v(t) \ln \frac{v(t)}{u(t)} - v(t) + u(t) \right] dt, \qquad (1)$$

when it is finite.

Some literature on recovering nonnegative solutions

- Maximum entropy regularization

$$\min_u \|Au - y\|^2 + \alpha f(u)$$

- Joint Kullback-Leibler regularization

$$\min_u d(y, Au) + \alpha d(u, u^*)$$

Computationally: nonlinear optimization problems.

Amato, Hughes, Engl, Landl, Eggermont, Anderssen, R,...

Expectation-Maximization algorithm for Poisson models

$$u_{k+1}(t) = u_k(t)\, A^* \frac{y}{Au^k}(t)$$

Advantages:

- it shapes the features of the solution in early iterations
- easy to compute

Disadvantages:

- slow algorithm
- very unstable numerically

A class of iterative methods

$$u_k \in \arg\min_{u \in X} \left\{ \frac{1}{2} \| Au - y \|^2 + cd(u, u_{k-1}) - \frac{1}{2} \| Au - Au_{k-1} \|^2 \right\},$$

where

- $d = D_R$ denotes the Bregman distance associated with a convex functional $R : X \to [0, +\infty]$
- c is some positive number.

Equivalently (when Y is a Hilbert space):

$$u_k \in \arg\min_{u} \left\{ \langle u, A^*(Au_{k-1} - y) \rangle + cd(u, u_{k-1}) \right\}.$$

Some methods of this type in the IP literature

- The Landweber method:

$$u_k = \text{argmin}_u \left\{ \|Au - y\|^2 + c\|u - u_{k-1}\|^2 - \|Au - Au_{k-1}\|^2 \right\},$$

where $c = 1$ if $\|A\| \leq 1$. Equivalently,

$$u_k = u_{k-1} + A^*(y - Au_{k-1}).$$

- Daubeschies, Defries, De Mol '04 (surrogate functionals):

$$u_k = \text{argmin}_u \left\{ \|Au - y\|^2 + \lambda R(u) + c\|u - u_{k-1}\|^2 - \|Au - Au_{k-1}\|^2 \right\}.$$

- Ramlau, Teschke '05: As above, for nonlinear operators.
- Schöpfer, Louis, Schuster '06 (method in Banach spaces)

$$J_X(x_{n+1}) = J_X(x_n) + \mu_n A^* J_Y(y - Ax_n),$$

where $A : X \to Y$, J_X and J_Y are duality mappings.

Some methods of this type in the IP literature

- linearized Bregman iteration - Osher et al. '08

$$u_{k+1} = \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \mu D_R(u, u_k) + \frac{1}{2c} \| u - (u_k - cA^*(Au_k - y)) \|^2 \right\},$$

$$p_{k+1} = p_k - \frac{1}{\mu c}(u_{k+1} - u_k) - \frac{1}{\mu} A^*(Au_k - y),$$

for some $\mu, c > 0$.
Here $D_R(u, u_k) = R(u) - R(u_k) - \langle p_k, u - u_k \rangle$.

Reformulation:

$$u_{k+1} = \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \langle Au, Au_k - y \rangle + \mu D_R(u, u_k) + \frac{1}{2c} \| u - u_k \|^2 \right\},$$

Convergence: Finite dimension, for smooth functionals R.

Related literature in finite dimensional optimization

$$u_k \in \arg\min_u \{\langle u, A^*(Au_{k-1} - y)\rangle + cd(u, u_{k-1})\},$$

with $d = D_f = KL$, f being the entropy: $f(u) = \sum_{j=1}^n u_j \ln u_j$.

Start with:

$$\min_{u \geq 0} g(u)$$

- Proximal point methods:

$$u_{k+1} = \text{argmin}_u g(u) + c_k d(u, u_k)$$

  Implicite iterative method

- Easier: Linearize the objective functional, i.e., $g(u) \sim g(u_k) + \nabla g(u_k)^t (u - u_k)$

$$u_{k+1} = \text{argmin}_u \nabla g(u_k)^t u + c_k d(u, u_k)$$

$$u_{k+1} = \mathrm{argmin}_u \nabla g(u_k)^t u + c_k d(u, u_k)$$

The first order optimality condition for this problem is

$$\nabla f(u_{k+1}) = \nabla f(u_k) - \frac{1}{c_k} \nabla g(u_k),$$

Since ∇f invertible,

$$u_{k+1} = (\nabla f)^{-1} (\nabla f(u_k) - \frac{1}{c_k} \nabla g(u_k))$$

that is

$$u_{k+1}^j = u_k^j e^{-\lambda_k \nabla g(u_k)^j}, \quad \lambda_k = 1/c_k$$

Iusem '94, '97

Convergence of the finite dimensional optimization method

- Steepest descent method for unconstrained optimization: $\min_u g(u)$

$$u_{k+1} = u_k - t_k \nabla g(u_k),$$

with $t_k > 0$ minimizing the one dimensional function $t \mapsto g(u_k - t\nabla g(u_k))$

- 'steepest descent method' for the constrained problem $\min_{u \geq 0} g(u)$:

$$u_{k+1} = (\nabla f)^{-1}(\nabla f(u_k) - \frac{1}{c_k} \nabla g(u_k))$$

More precisely:

$$u^j_{k+1} = u^j_k e^{-t_k \nabla g(u_k)^j},$$

with

$$t_k = \mathrm{argmin}_{t>0} g(u(t)), \quad u(t)^j = u^j_k e^{-t\nabla f(u_k)^j}$$

Other works on entropic projections methods: D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyre '15

Entropic projection method in infinite dimensional spaces

The iterative method we analyse:

$$u_k \in \arg\min_u \left\{ \langle Au, Au_{k-1} - y \rangle + cd(u, u_{k-1}) + \chi_j(u) \right\},$$

where

$$\chi_1(u) = \begin{cases} 0 & \text{if } \int_\Omega u(t)\ dt = 1, \\ +\infty & \text{else,} \end{cases}$$

and $\chi_0 \equiv 0$
(the original problem without integral constraint).

Properties of the entropy functionals

- $dom\,\partial f(u) = \{u \in L^\infty_+(\Omega), u \text{ is bounded away from zero}\}$. Moreover, $\partial f(u) = \{1 + \log u\}$.
- For all $u, v \in dom\, f$,

$$\|u - v\|_1^2 \leq \left(\frac{2}{3}\|v\|_1 + \frac{4}{3}\|u\|_1\right) d(v, u).$$

- The function $d(\cdot, u^*)$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$, $\forall\, u^* \in dom\, f$.
- The following sets are weakly compact in $L^1(\Omega)$:

$$\{x \in L^1(\Omega) : d(x, u) \leq C\}, \;\; \forall\, C > 0, \forall\, u \in dom\, f.$$

- The set $\partial d(\cdot, u^*)(u)$ is nonempty for $u^* \in dom\, f$ if and only if u belongs to $L^\infty_+(\Omega)$ and is bounded away from zero. Moreover, $\partial d(\cdot, u^*)(u) = \{\log u - \log u^*\}$.

Welldefinedness of the iterative scheme

Proposition: Let $\ell \in L^\infty(\Omega)$ and $v \in \operatorname{dom} \partial f$. Then the problem

$$\langle \ell, u \rangle + d(u, v) + \chi_j(u) \to \min_{u \in \operatorname{dom}(f)} \tag{2}$$

has a unique solution in the cases $j = 0$ and $j = 1$, respectively, given by

$$u_j = c_j v e^{-\ell}, \qquad c_j = \begin{cases} 1 & \text{if } j = 0, \\ \frac{1}{\int_\Omega v e^{-\ell} \, dt} & \text{if } j = 1, \end{cases} \tag{3}$$

which satisfies $u_j \in \operatorname{dom} \partial f$.

Proof: Rewrite the objective functional as

$$\begin{aligned}\langle \ell, u\rangle + d(u,v) + \chi_j(u) &= \int_\Omega \left[u(t)\ln\frac{u(t)}{v(t)} - v(t) + u(t) + u(t)\ell(t)\right] dt \\ &+ \chi_j(u) \\ &= \int_\Omega \left[u(t)\ln\frac{u(t)}{u_j(t)} + v(t) - u(t) + u(t)\ln c_j\right] dt \\ &+ \chi_j(u) \\ &= d(u,u_j) + \ln c_j\left(\int_\Omega u(t)\ dt - 1\right) + \chi_j(u) + C_j.\end{aligned}$$

Notice that

$$\ln c_j\left(\int_\Omega u(t)\ dt - 1\right) = 0.$$

Hence, the problem is equivalent to minimizing $d(u,u_j) + \chi_j(u)$, with u_j as unique solution: the entropic Bregman projection of u_j.

Forward operators and entropy

- $A : L^1(\Omega) \to Y$ is a linear and bounded operator, Y is a Hilbert space.
- 'Continuity' Assumption:

$$\|Au - Av\| \leq \gamma\sqrt{d(u,v)} \tag{4}$$

holds on $\mathrm{dom}(f + \chi_j)$ when $j = 0$ or $j = 1$, for some $\gamma > 0$.

  - $j = 1$: Assumption holds due to boundedness of A:

$$\|Au - Av\|^2 \leq \|A\|^2 \ \|u - v\|^2_{L^1(\Omega)} \leq 2\|A\|^2 \ d(u,v)$$

  - $j = 0$: We restrict the analysis to the class of operators A satisfying $cd(u,v) - \frac{1}{2}\|Au - Av\|^2 \geq 0$ for $u \in \mathrm{dom} f$ and $v \in \mathrm{dom}\ \partial f$.

- We define the nonlinear functional

$$D(u,v) = cd(u,v) - \frac{1}{2}\|Au - Av\|^2,$$

with $D(u,v) \geq 0$ for $u, v \in \mathrm{dom}(f + \chi_j)$, and $v \in \mathrm{dom}\ \partial f$.

The entropic projection method (EPM)

$$u_k \in \arg\min_u \left\{ \langle Au - y, Au_{k-1} - y \rangle + cd(u, u_{k-1}) + \chi_j(u) \right\}$$

Proposition: Let $u_0 \in \text{dom } \partial f$. Then the iterates of the entropic projection method are well-defined for $k \geq 1$, i.e. the above problem has a unique minimizer, given by $(\lambda = \frac{1}{c})$

$$u_k = u_{k-1} c_{k-1}^j e^{\lambda A^*(y - Au_{k-1})}, \qquad c_{k-1}^j = \begin{cases} 1 & \text{if } j = 0, \\ \frac{1}{\int_\Omega u_{k-1} e^{\lambda A^*(y - Au_{k-1})} \, dt} & \text{if } j = 1, \end{cases}$$

which further satisfies $u_k \in \text{dom } \partial f$.

Non-negativity: u_0 nonnegative $\Rightarrow$ u_k nonnegative, $\forall k \in N$.

Preparations for the convergence analysis

- The first-order optimality condition for the variational problem:

$$\ln u_k = \ln u_{k-1} + \ln c_{k-1}^j + \lambda A^*(y - Au_{k-1}),$$

where $\ln c_{k-1}^0 = 0$ and $\ln c_{k-1}^1$ - a Lagrange multiplier for the integral constraint.

- A proximal point method type:

$$u_k \in \arg\min_u \left\{ \frac{1}{2}\|Au - y\|^2 + \chi_j(u) + D(u, u_{k-1}) \right\},$$

$$D(u, u_{k-1}) = cd(u, u_{k-1}) - \frac{1}{2}\|Au - Au_{k-1}\|^2,$$

Difficulty: D is neither a metric distance nor even a Bregman distance, rather a weighted difference of Bregman distances.

Convergence analysis - exact data case

Proposition: If

- $A : L^1(\Omega) \to Y$ is a bounded linear operator which satisfies the 'continuity' condition
- $Au = y$ has a positive solution z verifying $\chi_j(z) = 0$ if $j = 1$
- $u_0 \in dom\,\partial f$ is an arbitrary starting element such that $1 + \log u_0 \in \mathcal{R}(A^*)$ and $\chi_j(u_0) = 0$ if $j = 1$,

  then the following statements are true:

  i) The residual $\|Au_k - y\|$ decreases when k increases;
  ii) The term $D(z, u_k)$ decreases when k increases;
  iii) The sequences $\{u_k\}_{k\in\mathbb{N}}$ generated by the entropic projection method converge weakly on subsequences in $L^1(\Omega)$ to solutions of the equation $Au = y$, with $\chi_j(u) = 0$ if $j = 1$.

Iterative methods for ill-posed problems have a typical behavior: The distance between the solution and the iterates decays initially, then it increases.

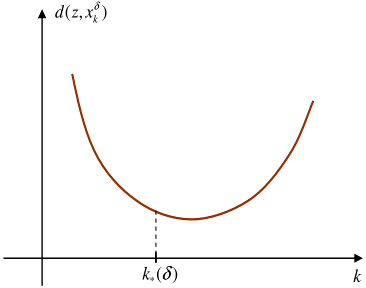
Iterative methods for ill-posed problems have a typical behavior: The distance between the solution and the iterates decays initially, then it increases.

Iterative methods for ill-posed problems have a typical behavior: The distance between the solution and the iterates decays initially, then it increases.

Engl, Hanke, Neubauer '96

Iterative methods for ill-posed problems have a typical behavior: The distance between the solution and the iterates decays initially, then it increases.

Engl, Hanke, Neubauer ‘96



Convergence analysis - Noisy data case
Discrepancy principle

Proposition: If

- $A : L^1(\Omega) \to Y$ is bounded and linear, satisfying the 'continuity' condition
- z is a positive solution of $Au = y$ with $\chi_j(z) = 0$ if $j = 1$.
- $y^\delta \in Y$ are noisy data satisfying $\|y - y^\delta\| \le \delta$, for some noise level δ
- $u_0 \in dom\,\partial f$ is an arbitrary starting element with the properties $1 + \log u_0 \in \mathcal{R}(A^*)$ and $\chi_j(u_0) = 0$ if $j = 1$
- the stopping index k_* is chosen such that

$$k_*(\delta) = \max\{k \in \mathbb{N} : \|Au_k - y^\delta\| \ge \sqrt{\tau}\delta\}, \;\; \tau > 1,$$

then

i) The residual $\|Au_k - y^\delta\|$ decreases when k increases and

$$\frac{1}{2}\|y^\delta - Au_{k+1}\|^2 + D(z, u_{k+1}) + D(u_{k+1}, u_k) \le \frac{\delta^2}{2} + D(z, u_k), \quad k \in \mathbb{N}.$$

ii) The index $k_*(\delta)$ is finite;
iii) There exists a weakly convergent subsequence of $(u_{k_*(\delta)})_\delta$ in $L^1(\Omega)$. If $(k_*(\delta))_\delta$ is unbounded, then each limit point is a solution of $Au = y$.

Convergence analysis - Noisy data case II
A priori rule

Proposition: If

- $A : L^1(\Omega) \to Y$ is bounded and linear, satisfying the 'continuity' condition
- z is a positive solution of $Au = y$ with $\chi_j(z) = 0$ if $j = 1$.
- $y^\delta \in Y$ are noisy data satisfying $\|y - y^\delta\| \leq \delta$, for some noise level δ
- $u_0 \in dom\, \partial f$ is an arbitrary starting element with the properties $1 + \log u_0 \in \mathcal{R}(A^*)$ and $\chi_j(u_0) = 0$ if $j = 1$.
- the stopping index k_* us chosen of order $1/\delta$,

then the sequence $(f(u_{k_*(\delta)}))_\delta$ is bounded and thus, there exists a subsequence of $(u_{k_*(\delta)})_\delta$ in $L^1(\Omega)$ which converges weakly to a solution of $Au = y$.

Error estimates - exact data case

Proposition: If

- $A : L^1(\Omega) \to Y$ is a bounded linear operator satisfying the 'continuity' condition
- z is a positive solution of $Au = y$ verifying $\chi_j(z) = 0$ if $j = 1$
- $u_0 \in dom\,\partial f$ be an arbitrary starting element with the properties $1 + \log u_0 \in \mathcal{R}(A^*)$ and $\chi_j(u_0) = 0$ if $j = 1$.

Additionally, let the following source condition hold:

$$1 + \log z \in \mathcal{R}(A^*).$$

Then one has

$$d(z, u_k) = O(1/k).$$

Moreover, $\|u_k - z\|_1 = O(1/\sqrt{k})$ if $j = 1$.

A sketch of the proof

We consider only the case $j = 0$, i.e. no constraints (similar arguments for the other case).

- Let $D^s(x, y) = D(x, y) + D(y, x)$.
- Let $\xi = 1 + \log z = \lambda A^* v$ for some $v \in Y$,
  $\xi_0 = 1 + \log u_0 = \lambda A^* w_0$ for some $w_0 \in Y$
- $$v_k = w_0 + \sum_{j=0}^{k-1} (y - Au_j), \quad k \geq 1$$
- The optimality condition: $\xi_k = \xi_{k-1} + \lambda A^*(y - Au_{k-1})$ implies

$$\xi_k = \xi_0 + \lambda A^* \Big(\sum_{j=0}^{k-1} (y - Au_j) \Big) = \lambda A^* v_k.$$

Burger, R, He '07 (similar technique)

$$
\begin{aligned}
D^s(u_k, z) &= c\langle \xi_k - \xi, u_k - z\rangle - \|Au_k - Az\|^2 \\
&= \langle A^* v_k - A^* v, u_k - z\rangle - \|Au_k - y\|^2 \\
&= \langle v_k - v, Au_k - y\rangle - \|Au_k - y\|^2 \\
&= \langle v_k - v, v_k - v_{k+1}\rangle - \|Au_k - y\|^2 \\
&= \frac{1}{2}\|v_k - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 + \frac{1}{2}\|v_{k+1} - v_k\|^2 - \|Au_k - y\|^2 \\
&= \frac{1}{2}\|v_k - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 - \frac{1}{2}\|Au_k - y\|^2.
\end{aligned}
$$

By writing the last inequality also for $k-1, k-2, \ldots, 1$, by summing up and by combining with monotonicity of $\{D(z, u_k)\}$, one obtains

$$
kD(z, u_k) \leq \sum_{j=1}^{k} D^s(u_j, z) \leq \frac{1}{2}\|v_1 - v\|^2 - \frac{1}{2}\|v_{k+1} - v\|^2 - \frac{1}{2}\sum_{j=1}^{k}\|Au_j - y\|^2
$$

and thus,

$$
d(z, u_k) \leq \frac{\lambda}{2k}\|v_1 - v\|^2.
$$

Error estimates - noisy data case

Proposition: If

- $A : L^1(\Omega) \to Y$ is a bounded linear operator satisfying the 'continuity' condition
- z is a positive solution of $Au = y$ verifying $\chi_j(z) = 0$ if $j = 1$
- $u_0 \in dom\,\partial f$ be an arbitrary starting element with the properties $1 + \log u_0 \in \mathcal{R}(A^*)$ and $\chi_j(u_0) = 0$ if $j = 1$.
- $y^\delta \in Y$ are noisy data satisfying $\|y - y^\delta\| \le \delta$, for some noise level δ

Additionally, let the following source condition hold:

$$1 + \log z \in \mathcal{R}(A^*),$$

and choose $k_*(\delta) \sim \frac{1}{\delta}$. Then one has

$$d(z, u_{k_*(\delta)}) = O(\delta).$$

Moreover, $\|u_{k_*(\delta)} - z\|_1 = O(\sqrt{\delta})$ if $j = 1$.

Further steps to be pursued

- Strong convergence in $L^1(\Omega)$.
- Convergence rates in the noisy data case (discrepancy principle).
- Numerical examples

Last, but not least:

Weaken the 'continuity' condition on the operator and/or find relevant problems fulfilling it and test the EPM numerically.

References

- U. Amato and W. Hughes, Maximum entropy regularization of Fredholm integral equations of the first kind, Inverse Problems 7 (1991) 793–803.
- M. Bachmayr. Iterative Total Variation Methods for Nonlinear Inverse Problems. Master Thesis, J.K. University, Linz, 2007.
- M. Burger, E. Resmerita, L.He, Error estimation for Bregman iterations and inverse scale space methods in image restoration. Computing 81 (2007) 109–135
- M. Burger and E. Resmerita, Iterative Regularization of Linear Ill-posed Problem by an Entropic Projection Method, preprint, 2016.
- I. Daubechies, M. Defrise and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math* 57 (2004) 1413–1457.
- P. Eggermont. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.* 24 (1993) 1557–1576.
- H. Engl and G. Landl. Convergence-rates for maximum-entropy

- A.N.Iusem, Steepest Descent Methods with Generalized Distances for Constrained Optimization. Acta Applicandae Mathematicae. 46(2) (1997) 225246.
- A. Kondor. Method of convergent weights An iterative procedure for solving Fredholm's integral equations of the first kind. *Nucl. Instr. Meth. Phys. Res.* 216 (1983) 177–181.
- H.N. Mülthei and B. Schorr. On an iterative method for a class of integral equations of the first kind. *Math. Methods Appl. Sci.* 9(2) (1987) 137–168.
- J-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyr. Iterative Bregman Projections for Regularized Transportation Problems. SIAM Journal on Scientific Computing, 37(2) (2015) A1111A1138.
- E. Resmerita and R.S. Anderssen. A joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. *Math. Meth. Appl. Sci.* DOI: 10.1002/mma.855 2007.
- F. Schoepfer, A.K. Louis and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems* 22 (2006) 311–329.