A New Non-Monotonic Algorithm for PET Image Reconstruction

Suvrit Sra, Dongmin Kim, Inderjit Dhillon, and Bernhard Schölkopf

***Abstract*—Maximizing some form of Poisson likelihood (either with or without penalization) is central to image reconstruction algorithms in emission tomography. In this paper we introduce NMML, a non-monotonic algorithm for maximum likelihood PET image reconstruction. NMML offers a simple and flexible procedure that also easily incorporates standard convex regularization for doing penalized likelihood estimation. A vast number image reconstruction algorithms have been developed for PET, and new ones continue to be designed. Among these, methods based on the expectation maximization (EM) and ordered-subsets (OS) framework seem to have enjoyed the greatest popularity. Our method NMML differs fundamentally from methods based on EM: i) it does not depend on the concept of *optimization transfer* (or surrogate functions); and ii) it is a rapidly converging *non-monotonic* descent procedure. The greatest strengths of NMML, however, are its simplicity, efficiency, and scalability, which make it especially attractive for tomographic reconstruction. We provide a theoretical analysis NMML, and empirically observe it to outperform standard EM based methods, sometimes by orders of magnitude. NMML seamlessly allows integreation of penalties (regularizers) in the likelihood. This ability can prove to be crucial, especially because with the rapidly rising importance of combined PET/MR scanners, one will want to include more "prior" knowledge into the reconstruction.**

***Index Terms*—Non-monotonic maximum likelihood, ordered subsets expectation maximization (OS-EM), emission tomography, transmission tomography, convex optimization, penalized likelihood.**

I. INTRODUCTION

In this paper we present a new algorithm called Non-Monotonic Maximum Likelihood (NMML) for computing both penalized and non-penalized ML solutions with a focus on image reconstruction for PET. We note that NMML can be easily extended to other inverse problems, e.g., transmission tomography, image restoration, etc. [1].

The distinguishing feature of NMML is its non-monotonic optimization of the likelihood, which allows it to take more aggressive steps, and therefore to achieve rapid convergence. More importantly, we prove convergence of NMML, despite its non-monotonicity [1], which differentiates NMML from most other heuristic non-monotonic approaches.

Our goal while designing NMML was to satisfy the following stringent requirements:

1) rapid convergence; ideally orders of magnitude faster than MLEM, and at least as fast as OSEM,
2) ease of implementation and use; minimum amount of parameter tuning on part of the user,
3) ability to rigorously and *easily* handle regularization, and
4) theoretical guarantees on convergence.

These desiderata are very challenging to simultaneously fulfill. Often methods having theoretical guarantees on convergence simply do not compete *empirically* with OSEM, or lack implementation ease. However, OSEM does have one major drawback, namely, its inability to rigorously handle regularization (though several workarounds such as using the modified EM technique of [2], or Green's OSL heuristic [3] have been suggested). Regularization ability is also important due to another recent development: combined PET/MR scanners; here image priors obtained from MR can be used to improve the PET image reconstruction via appropriate regularizers.

Our initial results with NMML are quite encouraging. We see NMML to be rapidly convergent, flexible, and competitive to OSEM. For well-conditioned system matrices NMML can even be orders of magnitude better than OSEM (Fig. 1)! In more realistic settings, NMML is competitive to OSEM, and for noisy-data, NMML yields good results where OSEM fails. Our results are thus a "proof-of-concept" and feedback from the medical imaging community will be valuable in transferring our algorithm into a clinical setting.

II. PET RECONSTRUCTION

We follow the standard PET setup [4], wherein the *sinogram* counts for detector i are modeled by

$$y_i \sim \text{Poisson}(c_i[\boldsymbol{A}\boldsymbol{x}]_i + r_i), \tag{1}$$

where $\boldsymbol{A}$ is the system matrix, $\boldsymbol{x}$ the image that we wish to reconstruct, r_i the average accidental coincindences, and c_i the detector calibration factors [5]. For simplicity, we will present our method with $r_i = 0$ and $c_i = 1$, noting that our approach can handle the general case too if *convexity* is ensured.

Maximizing the log-likelihood corresponding to (1), with the (natural) constraint $\boldsymbol{x} \geq 0$ is equivalent to solving

$$\min_{\boldsymbol{x}\geq 0} \quad f(\boldsymbol{x}) = \sum_i -y_i \log[\boldsymbol{A}\boldsymbol{x}]_i + [\boldsymbol{A}\boldsymbol{x}]_i. \tag{2}$$

Often, to prevent overfitting of noise, to smoothen the image, or to incorporate prior knowledge, one adds a *regularization function* to (2), which results in the problem

$$\min_{\boldsymbol{x}\geq 0} \quad h(\boldsymbol{x}) = f(\boldsymbol{x}) + \beta R(\boldsymbol{x}), \tag{3}$$

where $\beta > 0$ is a penalty parameter and $R(\boldsymbol{x})$ is a regularizing function. Several choices of $R(\boldsymbol{x})$ have been studied in the literature [5], e.g., $R(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|^2$ is the traditional energy penalty, while $R(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{C}\boldsymbol{x}\|^2$ where $\boldsymbol{C}$ is a finite-differencing matrix provides a first-order roughness penalty. For our algorithm, we only make the requirement that $R(\boldsymbol{x})$ be a differentiable convex function of $\boldsymbol{x}$, and no additional effort (derivation of surrogates etc.) is needed for using it.

SS, BS are with the Max Planck Institute for Biological Cybernetics.
DK, ID are with the University of Texas at Austin.

9781-4244-3962-1/09/$25.00 ©2009 IEEE

A. The NMML Algorithm

We present simplified pseudo-code for NMML as Algorithm 1, though we must omit the theoretical analysis (including convergence) due to lack of space—the longer version of this paper with contain these details. NMML extends the successful *unconstrained* optimization method of Barzilai-Borwein (BB) [6] to the *constrained* case. Our extension is non-trivial because naive approaches to using the BB technique for constrained problems leads to divergent or oscillatory behavior. For details on our extensions that permit us to deal with constrained problems, please see [1], [7].

Broadly viewed, the BB method proposes a special choice of step-size for a gradient-projection method, obviating the for expensive procedures like *line-search*. This results in a much simpler but still rapidly converging algorithm. For the constrained case, the computation of these BB step-sizes must be modified carefully using the notion of a *fixed set*, i.e., the variables that should *not* participate in optimization at the current step; Algorithm 1 shows the details.

Other than the fixed set I_+, the most important computation in Algorithm 1 is the gradient

$$\nabla h(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) + \beta \nabla R(\boldsymbol{x}) = \boldsymbol{A}^T(\boldsymbol{1} - \hat{\boldsymbol{y}}) + \beta \nabla R(\boldsymbol{x}), \quad (4)$$

where $\hat{\boldsymbol{y}}$ is given by the elementwise division $[y_i/[\boldsymbol{Ax}]_i]$ (using the standard assumption that $[\boldsymbol{Ax}]_i \neq 0$). Thus, computing (4) requires one *forward projection*, one *back projection*, and evaluation of $\nabla R(\boldsymbol{x})$, whereby the total computational cost remains very low.

Algorithm 1: Non-monotonic maximum likelihood

```
Input: $\boldsymbol{y}$: sinogram, $h$: funct. to optimize
Output: $x$ reconstructed image, i.e.,solution to (3)
Initialize $k \leftarrow 1$, $\boldsymbol{x}^0 \in \mathbb{R}^n_{++}$,   $\boldsymbol{g}^0 = \nabla h(\boldsymbol{x}^0)$;
Set $\alpha_0$, s.t. $h(\boldsymbol{x}^1) < h(\boldsymbol{x}^0)$ for $\boldsymbol{x}^1 \leftarrow P_+(\boldsymbol{x}^0 - \alpha_0 \boldsymbol{g}^0)$;
repeat
    $\boldsymbol{g}^k = Z_+\big(\nabla h(\boldsymbol{x}^k)\big)$;
    Compute fixed set $I_+ = \big\{i \big| x_i^k = 0, [\boldsymbol{g}^k]_i > 0\big\}$;
    $\Delta\boldsymbol{x} \leftarrow Z_+(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})$   {Zero out variables in $I_+$};
    $\Delta\boldsymbol{g} \leftarrow Z_+(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})$   {Zero out gradient};
    $\alpha_k \leftarrow \frac{\langle \Delta\boldsymbol{x}, \Delta\boldsymbol{x} \rangle}{\langle \Delta\boldsymbol{x}, \Delta\boldsymbol{g} \rangle}$   {Compute BB step};
    $\boldsymbol{x}^{k+1} \leftarrow P_+(\boldsymbol{x}^k - \alpha_k \boldsymbol{g}^k)$;
    $k \leftarrow k + 1$;
until Stopping Criteria are met ;
```

B. Implementation Details for PET

The individual steps of the method described above are displayed in Algorithm 1, and with a little care they can be implemented highly efficiently. Note that the gradient computation

$$[\nabla h(\boldsymbol{x})]_j = \sum_i a_{ij} - \sum_i \frac{y_i a_{ij}}{[\boldsymbol{Ax}]_i} + \beta[\nabla R(\boldsymbol{x})]_j,$$

can be written as

$$\nabla h(\boldsymbol{x}) = \boldsymbol{A}^T(\boldsymbol{1} - \hat{\boldsymbol{y}}) + \beta \nabla R(\boldsymbol{x}), \quad (5)$$

where $\hat{\boldsymbol{y}}$ is given by the elementwise division $[y_i/[\boldsymbol{Ax}]_i]$. In the language of PET image reconstruction (5) requires one *forward projection* and one *back projection*—in matrix terminology, it requires just two matrix-vector BLAS2 operations, which can be computed very efficiently, or even parallelized easily if needed. Finally, care must be taken to avoid any division by zero while computing the gradient.

C. Convergence

One can show the following convergence theorem—see [1] for details.

Theorem 1 (ϵ-optimal Convergence). *If h^* denotes the optimal objective value for (3), and α_k as given by Algorithm 1 is bounded above, then there exists a constant $\epsilon > 0$, such that*

$$\lim_{k\to\infty} \bar{h}^k - h^* < \epsilon,$$

where $\bar{h}^k$ is the best objective so far, i.e., $\bar{h}^k = \min\{\bar{h}^{k-1}, h(\boldsymbol{x}^k)\}$.

III. EXPERIMENTS

We show two main experiments, both of which used 256×256 images, and 256×192 sinograms, so that $\boldsymbol{A}$ was 49152×65536 in size. For OSEM we used Fessler's powerful IRT software [8]; NMML was also implemented in MATLAB.

Our first set of experiments shows results with a *well-conditioned* system matrix $\boldsymbol{A}$, and is included merely to indicate the potential speedup NMML can offer. The plots in Fig. 1 show convergence of the objectives for NMML and OSEM, with increasing photon counts from left to right. NMML yields objectives up to 2 orders of magnitude better than OSEM (tried with 8, 16, 32 subsets) in the same time.

The second set of experiments is on simulated noisy (to very noisy) PET sinograms (first col. of Fig. 2), and a corresponding *ill-conditioned* system matrix $\boldsymbol{A}$. For this difficult setup, NMML was run with a first-order smoothness penalty as the regularizer. We tried to run OSEM with the OSL heuristic of [3], however, could not obtain any meaningful results with it. We thus used the OS version of De Pierro's technique [2], OSDP, taken from the IRT. We tuned the value of β used for the regularizer, and report results for the most favorable values (used for both OSDP and NMML). The second, third, and last columns of Fig. 2 report the results produced by plain OSEM, OSDP, and NMML. It is easy to see that for the high-noise cases (top two rows), only NMML yields something slightly meaningful. For decreasing noise, OSEM still overfits, while OSDP and NMML yield comparable results. We mention that NMML much faster (3 to 4 times) than OSDP (detailed timing results will be shown in the longer version of the paper).

Our experimental results are encouraging, and indicate that NMML has potential as a rigorous convex optimization method for PET image reconstruction, especially for situations where regularization is important. Its simplicity should make it an easy candidate for adoption in practice too, and running NMML on real clinical data is a part of our future work.

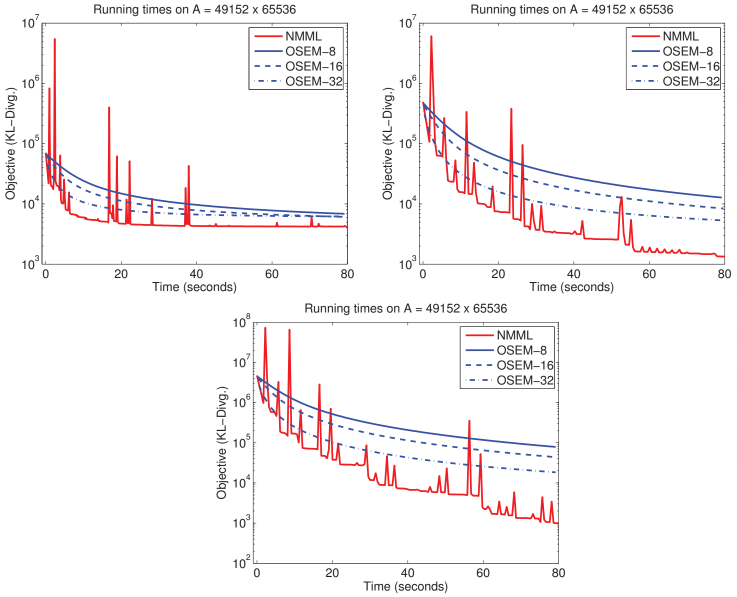

Fig. 1. Convergence of NMML compared to OSEM (8, 16, 32 subsets), both without *regularization*. From left to right (and top-down) data has increasing photon counts. Note that the y-axis in on a logarithmic scale.

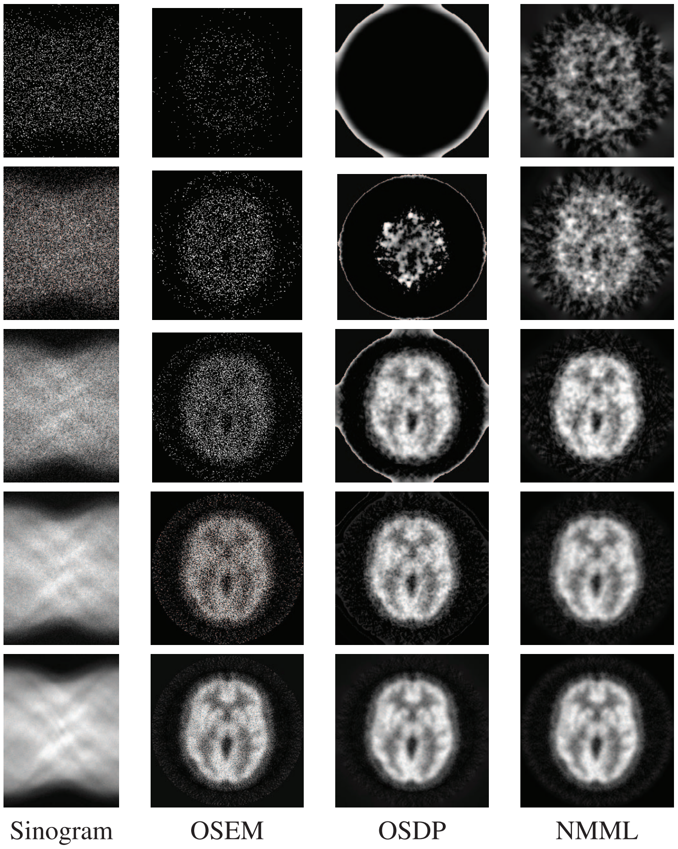

Fig. 2. Regularized reconstruction for sinograms with varying noise levels. Both OSEM and OSDP fail totally for the high-noise (rows 1 & 2). With decreasing noise (top–bot), OSDP and NMML perform comparably; though NMML runs several times faster (timing not reported due to lack of space).

REFERENCES

[1] S. Sra, D. Kim, and B. Schölkopf, “Non-monotonic Poisson Likelihood Maximization,” Tech. Rep. 170, Max Planck Institute for Biological Cybernetics, Jun. 2008.

[2] A. R. De Pierro, “A modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography,” *IEEE Tran. Med. Imag.*, vol. 14, no. 1, pp. 132–137, 1995.

[3] P. J. Green, “Bayesian reconstructions from emission tomography data using a modified EM algorithm,” *IEEE TMI*, vol. 9, no. 1, 1990.

[4] L. A. Shepp and Y. Vardi, “Maximum likelihood reconstruction for emission tomography,” *IEEE Tran. Medical Imaging*, vol. 1, pp. 113–122, 1982.

[5] J. Fessler, “Image reconstruction: Algorithms and Analysis,” Under preparation, 2009.

[6] J. Barzilai and J. M. Borwein, “Two-Point Step Size Gradient Methods,” *IMA J. Num. Analy.*, vol. 8, no. 1, 1988.

[7] D. Kim, S. Sra, and I. S. Dhillon, “A Non-monotonic Gradient Projection Method for the Non-negative Least Squares Problem,” Tech. Rep., Univ. of Texas at Austin, June 2008.

[8] J. Fessler, “ASPIRE & IRT Software Toolkits,” http://www.eecs.umich.edu/ fessler/, 2009.