A New, Fast, Relaxation-Free, Convergent Ordered-Subset Algorithm for Emission Tomography

Ing-Tsung Hsiao[1], Anand Rangarajan[3], Parmeshwar Khurd[2], Gene Gindi[2]
[1]School of Medical Technology,
Chang Gung University, Tao-Yuan 333, Taiwan
[2]Departments of Electrical & Computer Engineering and Radiology,
SUNY Stony Brook, Stony Brook, NY 11784, USA
[3]Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA

Abstract— **In this paper, we propose a fast, convergent OS-type (ordered-subset) reconstruction algorithm for emission tomography (ET) by taking into account the Hessian information in the ML Poisson objective. Most importantly, our proposed algorithm does not have a relaxation parameter and it is fundamentally not based on EM-ML algorithm in ET . Our new algorithm is based on an expansion of the ML objective using a second order Taylor series approximation w.r.t. the projection of the source distribution. Defining the projection of the source as an independent variable, we construct a new objective function in terms of the source distribution *and* the projection. This new objective function contains the Hessian information of the likelihood. After using a separable surrogate transformation of the new Hessian-based objective, we derive an ordered subsets, positivity preserving algorithm which is guaranteed to asymptotically reach the maximum of the original ET likelihood. Preliminary results show that this new algorithm is faster than our previous COSEM algorithm after some initial iterations and compatible with RAMLA. However, in contrast to RAMLA, and similar to COSEM, the new algorithm does not require any user-specified relaxation parameters.**

I. Introduction

Emission tomography in nuclear medicine, including PET and SPECT, is a useful diagnostic tool for investigating functions in an organ of interest in-vivo. To obtain a tomographic image of the injected radiotracer distribution in the body, one needs to apply a recontruction algorithm to collected data. Statistical reconstructin methods are capable of modelling photon noise and system phsyics more accurately than the traditional FBP method, and thus have drawn much attention in recent years. However, statistical reconstruction has the drawback of being slow when used for clinical studies as comparied to the FBP. To improve the speed of the statistical reconstruction, many fast reconstruction algorithms have been proposed in the past few years. Among them, the OS-type (ordered subsets) algorithms are the most popular approaches since the introduction of the OSEM algorithm in 1994[**?**].

We can rougly classify the OS-type algorithms for maximum likelihood-based (ML) reconstruction into three broad categories; i) heuristic OS-EM type algorithms which are not provably convergent, ii) convergent ordered subsets algorithms requiring a user-specified relaxation schedule and iii) convergent OS incremental EM type algorithms (COSEM). Our previous work on fast, provably convergent ML-EM algorithms is based on the third (COSEM) category above. Since EM algorithms (OS, incremental or otherwise) typically do not take into account the Hessian information in the ML objective, we now embark upon a new, fourth category of provably convergent, fast ordered subsets algorithms; iv) fast, convergent OS Hessian-based algorithms which do not have a relaxation parameter and which are fundamentally not based on EM.

To derive our new method, we expand the ML objective using a second order Taylor series approximation w.r.t. the projection of the source distribution. Defining the projection of the source as an independent variable, we construct a new objective function in terms of the source distribution *and* the projection. This new objective function contains the Hessian information of the likelihood. After using a separable surrogate transformation of the new Hessian-based objective, we derive an ordered subsets, positivity preserving algorithm which is guaranteed to asymptotically reach the maximum of the original ET likelihood.

This paper is to introduce and access the performance of the proposed algorithm. Section II we start on introducing the negative Poisson likelihood function, and then by the change of variables under the use of second order Taylor approximation, the new algorithm is derived step-by-step. In Section III, we run some empirical simulations to illustrate the speed of the proposed algorithm in comparison to ML-EM, COSEM and RAMLA. Section IV contains the discussion and the conclusion.

II. Theory

A. A new SP objective function for emission tomography

The emission tomography negative log-likelihood objective function is

$$E_{\rm ML}(\mathbf{f}) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_i g_i \log \sum_j \mathcal{H}_{ij} f_j. \qquad (1)$$

We would like to transform this objective in a manner similar to Fessler's recent OS-SPS work. In this previous work, a second order Taylor-series expansion is used to construct a surrogate for E_{ML}. In the present work, we also use a second order Taylor series expansion but instead of constructing a surrogate, we construct a new complete data SP objective function. This leads directly to a new C-SP algorithm which is similar in spirit to the EM algorithm. This approach also naturally allows us to construct COS-SP and ECOS-SP algorithms.

B. Change of variables

We wish to introduce a new variable $\{\sigma_i\}$ which will subsequently be identified with $\{\sum_j \mathcal{H}_{ij} f_j\}$. To achieve this, we take a second order Taylor series approximation of (1) at the location $\{\sigma_i = \sum_j \mathcal{H}_{ij} f_j\}$

$$E_2(\mathbf{f}, \sigma) = \sum_i \{\sigma_i - g_i \log \sigma_i + (1 - \frac{g_i}{\sigma_i})(\sum_j \mathcal{H}_{ij} f_j - \sigma_i) + \frac{g_i}{2\sigma_i^2}(\sum_j \mathcal{H}_{ij} f_j - \sigma_i)^2\}. \qquad (2)$$

The second order Taylor series expansion in (2) has the property that at $\{\sigma_i = \sum_j \mathcal{H}_{ij} f_j\}$, $E_2(\mathbf{f}, \sigma) = E_{\rm ML}(\mathbf{f})$. We wish to use (2) as the kernel of a new objective function which is to be minimized w.r.t. both $\mathbf{f}$ and σ. Since (2) does not necessarily have a minimum at $\{\sigma_i = \sum_j \mathcal{H}_{ij} f_j\}$, we augment (2) to obtain a new objective function

$$E_3(\mathbf{f}, \sigma) = \sum_i [\sigma_i - g_i \log \sigma_i + (1 - \frac{g_i}{\sigma_i})(\sum_j \mathcal{H}_{ij} f_j - \sigma_i) + \frac{g_i}{2\sigma_i^2}(\sum_j \mathcal{H}_{ij} f_j - \sigma_i)^2 + c_i \frac{(\sum_j \mathcal{H}_{ij} f_j - \sigma_i)^2}{\sigma_i \sum_j \mathcal{H}_{ij} f_j}]. \qquad (3)$$

There are two ways of choosing $\{c_i\}$. We briefly describe both approaches.

C. Choosing $\{c_i\}$ to ensure a global minimum at $\{\sigma_i = \sum_j \mathcal{H}_{ij} f_j\}$

In the first approach, we pick $\{c_i\}$ such that $E_3(\mathbf{f}, \sigma)$ has a global minimum at $\{\sigma_i = \sum_j \mathcal{H}_{ij} f_j\}$. This is NOT equivalent to convexity of the objective function w.r.t. σ. The chosen criterion is $\frac{\partial E_3}{\partial \sigma_i} \geq 0$ for $\sigma_i > \sum_j \mathcal{H}_{ij} f_j$. It can be shown that the resulted c_i only depends on g_i and not on $\sum_j \mathcal{H}_{ij} f_j$ as: $g_i(3 - 2\sqrt{2}) \leq c_i \leq g_i(3 - 2\sqrt{2})$. Since a lower value of $\{c_i\}$ is preferable, we choose $c_i = g_i(3 - 2\sqrt{2})$ as the optimum solution that guarantees that $\frac{\partial E_3}{\partial \sigma_i} \geq 0$ for $\sigma_i > \sum_j \mathcal{H}_{ij} f_j$.

D. Choosing c to ensure that the modified objective function is convex

In the second approach, $\{c_i\}$ is chosen such that $E_3(\mathbf{f}, \sigma)$ in (3) is a convex objective function w.r.t. σ.

We require that $\frac{\partial^2 E_3}{\partial \sigma_i^2} \geq 0$ for $\sigma_i > 0$. It can be shown that the resulted c_i is in the range by $g_i(3 - 2\sqrt{2}) \leq c_i \leq g_i(3 - 2\sqrt{2})$. We choose $c_i = g_i(2 - \sqrt{3})$. For this setting of c_i, $\frac{\partial^2 E_3}{\partial \sigma_i^2} = 0$ for $\sigma_i = \sqrt{3} \sum_j \mathcal{H}_{ij} f_j$ and nowhere else. The payoff in choosing the $\frac{(x-y)^2}{xy}$ form of the additional term is that c_i only depends on g_i and not on $\sum_j \mathcal{H}_{ij} f_j$.

E. Objective function w.r.t. f

We now examine the objective function as a function of $\mathbf{f}$. Since the new objective consists of terms such as $(\sum_j \mathcal{H}_{ij} f_j - \sigma_i)^2$ and $\frac{1}{\sum_j \mathcal{H}_{ij} f_j}$, we need to perform yet another change of variables so that we can update all $\mathbf{f}$ in parallel (while imposing positivity constraints). The objective function w.r.t. $\mathbf{f}$ alone is

$$E_{\rm PS}(\mathbf{f}) = \sum_i [(1 - \frac{g_i}{\sigma_i}) \sum_j \mathcal{H}_{ij} f_j + \frac{g_i}{2\sigma_i^2}\{(\sum_j \mathcal{H}_{ij} f_j)^2 - 2\sum_j \mathcal{H}_{ij} f_j \sigma_i\} + c_i(\frac{\sum_j \mathcal{H}_{ij} f_j}{\sigma_i} + \frac{\sigma_i}{\sum_j \mathcal{H}_{ij} f_j} - 2)]. \qquad (4)$$

The two terms in (4) which act as obstacles to a fully independent, parallel update of all $\mathbf{f}$ are the terms involving $(\sum_j \mathcal{H}_{ij} f_j)^2$ and $\frac{1}{\sum_j \mathcal{H}_{ij} f_j}$. We now detail two change of variables transformations to overcome this problem:

$$(\sum_j \mathcal{H}_{ij} f_j)^2 = \min_\rho \sum_j \frac{(\mathcal{H}_{ij} f_j)^2}{\rho_{ij}} \qquad (5)$$

subject to the constraint $\sum_j \rho_{ij} = 1$. Transforming this into a constrained optimization problem, we get

$$E_{\rm sumsq}(f, \rho, \mu) = \sum_j \frac{(\mathcal{H}_{ij} f_j)^2}{\rho_{ij}} + \mu_i(\sum_j \rho_{ij} - 1). \qquad (6)$$

When we minimize (6) w.r.t. ρ subject to the simplex constraint $\sum_j \rho_{ij} = 1$, we get

$$(\sum_j \mathcal{H}_{ij} f_j)^2 = \min_\rho \sum_j \frac{(\mathcal{H}_{ij} f_j)^2}{\rho_{ij}} + \mu_i(\sum_j \rho_{ij} - 1). \qquad (7)$$

The solution for ρ at the minimum is

$$\rho_{ij} = \frac{\mathcal{H}_{ij} f_j}{\sum_j \mathcal{H}_{ij} f_j}. \qquad (8)$$

Similarly, we can transform terms containing $\frac{1}{\sum_j \mathcal{H}_{ij} f_j}$. This can also be transformed into a constrained optimization problem via

$$\frac{1}{\sum_j \mathcal{H}_{ij} f_j} = \min_{\tau} \sum_j \frac{\tau_{ij}^2}{\mathcal{H}_{ij} f_j} + \nu_i (\sum_j \tau_{ij} - 1) \quad (9)$$

provided we satisfy the simplex constraint $\sum_j \tau_{ij} = 1$. The solution for τ at the minimum is

$$\tau_{ij} = \frac{\mathcal{H}_{ij} f_j}{\sum_j \mathcal{H}_{ij} f_j}. \quad (10)$$

Considering ρ and τ to be legitimate independent variables, we can write the transformed final objective function as

$$\begin{aligned} &E_{\text{final}}(\mathbf{f}, \sigma, \rho, \tau, \mu, \nu) = \\ &\sum_i \Big[\sigma_i - g_i \log \sigma_i + (1 - \frac{g_i}{\sigma_i})(\sum_j \mathcal{H}_{ij} f_j - \sigma_i) \\ &+ \frac{g_i}{2\sigma_i^2} \{ \sum_j \frac{(\mathcal{H}_{ij} f_j)^2}{\rho_{ij}} + \mu_i (\sum_j \rho_{ij} - 1) - 2 \sum_j \mathcal{H}_{ij} f_j \sigma_i + \sigma_i^2 \} \\ &+ c_i \{ \sum_j \frac{\tau_{ij}^2 \sigma_i}{\mathcal{H}_{ij} f_j} + \nu_i \sigma_i (\sum_j \tau_{ij} - 1) + \frac{\sum_j \mathcal{H}_{ij} f_j}{\sigma_i} - 2 \} \Big] \end{aligned} \quad (11)$$

Grouping terms in (11) which depend on $\mathbf{f}$, we get

$$E_j(f_j) = \sum_i \Big[\frac{c_i \tau_{ij}^2 \sigma_i}{\mathcal{H}_{ij} f_j} - \{ \frac{2g_i - c_i}{\sigma_i} - 1 \} \mathcal{H}_{ij} f_j + \frac{g_i \mathcal{H}_{ij}^2 f_j^2}{2\sigma_i^2 \rho_{ij}} \Big]. \quad (12)$$

This objective function is convex w.r.t. f_j and hence has a unique minimum. Taking the derivative w.r.t. f_j and setting the result to zero, we get

$$\begin{aligned} \frac{\partial E_j}{\partial f_j} = 0 \Rightarrow &- \sum_i \frac{c_i \tau_{ij}^2 \sigma_i}{\mathcal{H}_{ij} f_j^2} - \sum_i \{ \frac{2g_i - c_i}{\sigma_i} - 1 \} \mathcal{H}_{ij} \\ &+ \sum_i \frac{g_i \mathcal{H}_{ij}^2 f_j}{\sigma_i^2 \rho_{ij}} = 0. \end{aligned} \quad (13)$$

Equation (13) can be simplified into the following cubic equation form:

$$f_j^3 + a_2 f_j^2 + a_0 = 0 \quad (14)$$

where

$$a_2 \stackrel{\text{def}}{=} \frac{-\sum_i \{ \frac{2g_i - c_i}{\sigma_i} - 1 \} \mathcal{H}_{ij}}{\sum_i \frac{g_i \mathcal{H}_{ij}^2}{\sigma_i^2 \rho_{ij}}}, \text{ and } a_0 \stackrel{\text{def}}{=} \frac{-\sum_i \frac{c_i \tau_{ij}^2 \sigma_i}{\mathcal{H}_{ij}}}{\sum_i \frac{g_i \mathcal{H}_{ij}^2}{\sigma_i^2 \rho_{ij}}}. \quad (15)$$

If an OS update is used, one can show that the coefficients a_2 and a_0 in (15) can be further simplified as

$$a_2 = \frac{D_j - (2-c) \sum_l LB_j^{(k,l)}}{\sum_l \frac{1}{f_j^{(k,l)}} LB_j^{(k,l)}}, \text{ and} \quad (16)$$

$$a_o = \frac{-c \sum_l (f_j^{(k,l)})^2 LB_j^{(k,l)}}{\sum_l \frac{1}{f_j^{(k,l)}} LB_j^{(k,l)}}. \quad (17)$$

where the sensitivity is $D_j \stackrel{\text{def}}{=} \sum_i \mathcal{H}_{ij}$, and the limited backprojection $LB_j^{(k,l)} \stackrel{\text{def}}{=} \sum_{i \in S_l} \frac{g_i \mathcal{H}_{ij}}{\sum_n \mathcal{H}_{in} f_n^{(k,l)}}$. Since $c_i = c g_i$ where $c = 2 - \sqrt{3}$ or $c = 3 - 2\sqrt{2}$ depending on which solution is used, in order to compute a_2 and a_0, we only require the limited backprojection $LB_j^{(k,l)}$ which is exactly the same computational requirement as in the EM algorithm.

F. Real solutions to the cubic

Define $Q \stackrel{\text{def}}{=} -\frac{a_2^2}{9}$ and $R \stackrel{\text{def}}{=} -\frac{a_0}{2} - \frac{a_2^3}{27}$. Then the original cubic is transformed to04

$$(f_j + \frac{a_2}{3})^3 + 3Q(f_j + \frac{a_2}{3}) - 2R = 0. \quad (18)$$

The cubic discriminant is defined as

$$D \stackrel{\text{def}}{=} Q^3 + R^2 = \frac{a_0}{2}(\frac{a_0}{2} + \frac{a_2^3}{27}). \quad (19)$$

We distinguish between the cases $D > 0$ and $D \leq 0$. If $D > 0$, there is only one real solution to the cubic equation. If $D \leq 0$, multiple real solutions are present. If $D = 0$ (rare), all solutions are real and at least two are equal.

Case 1: $D > 0$

$$\hat{f}_j = -\frac{a_2}{3} + (R + \sqrt{D})^{\frac{1}{3}} + (R - \sqrt{D})^{\frac{1}{3}}. \quad (20)$$

Case 2: $D < 0$. Define

$$\theta \stackrel{\text{def}}{=} \cos^{-1}(\frac{R}{\sqrt{-Q^3}}). \quad (21)$$

Then the three real roots are

$$\hat{f}_j^{(1)} = -\frac{a_2}{3} + 2\sqrt{-Q} \cos(\frac{\theta}{3}), \quad (22)$$

$$\hat{f}_j^{(2)} = -\frac{a_2}{3} + 2\sqrt{-Q} \cos(\frac{\theta + 2\pi}{3}), \quad (23)$$

$$\hat{f}_j^{(3)} = -\frac{a_2}{3} + 2\sqrt{-Q} \cos(\frac{\theta + 4\pi}{3}). \quad (24)$$

We know *a priori* that only one real root can be positive. Regardless of whether $D > 0$ or $D < 0$, the one real, positive root is picked as the update for f_j.

III. Results

In this section, we anecdotally explore 2D reconstructions using the ML-EM, COSEM, RAMLA, and the proposed algorithms, and illustrate the speed-enhancing property by displaying the log-likelihood plots vs. iteration number.

To test the proposed algorithm, a noisy sinogram using a 2D 64×64 phantom is generated. The phantom consists of a disk background, two hot lesions and two cold lesions with contrast ratio of 1:4:8 (cold:background:hot). The projection data had dimensions of 64 angles by 96 detector bins. Poisson noise of 300K counts is simulated as well as uniform attenuation of water, and no other physical or geometrical blurring effects or background events are modeled here.

The sinogram is then reconstructed using the above mentioned algorithms, with the same subset of 4 except for ML-EM reconstruction. All reconstructions used a constant image as the initial estimate. The reconstructed images for each algorithm at 30th iteration are shown in Fig.1.

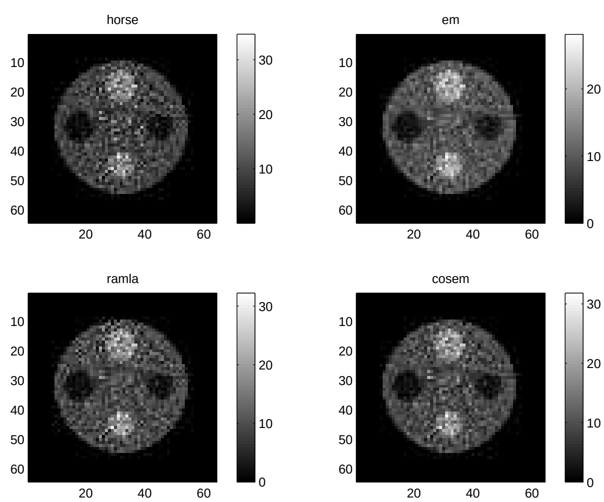

Fig. 1. 30 itertaions of all reconstructions.

We also plot the log-likelihood vs. iteration of each reconstruction in Fig.2 for ML-EM ('+'), COSEM('△'), RAMLA ('○'), and the proposed algorithm (HORSE?). As expected, the reconstructions of COSEM and RAMLA show relative order-of-magnitude acceleration over the ML-EM reconstruction, while COS-SP is slow in the beginning, but gets faster than EM (after 3th iteration), and than COSEM (after 4 iterations), and compatible with RAMLA (after 6 iterations).

IV. Discussion

We have proposed a new, provably convergent, OS-type and Hessian-based algorithm which does not have a relaxation parameter and which is fundamentally not based on EM. Preliminary results show that this new algorithm is faster than our previous COSEM algorithm

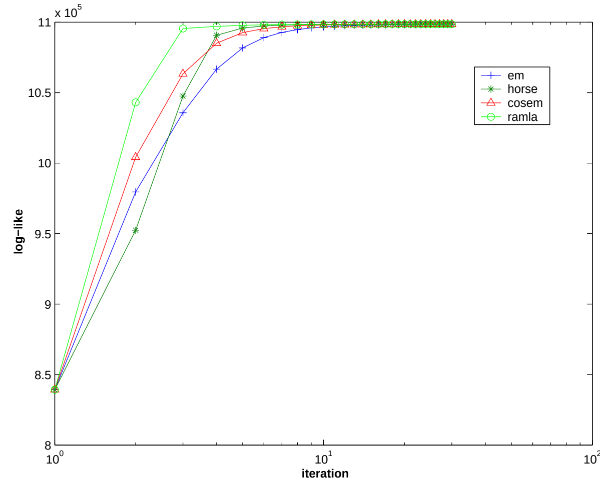

Fig. 2. The log-likelihood plots of all reconstructions. Every one has same IC (constant), and thus same log-likelihood in the first iteration (should be the "0-th" iteration.) Here, 4 subsets are used. For HORSE, $c = 2 - \sqrt{3} = 0.2679$. In the beginning, HORSE is slower than all (even EM), but got faster later.

after some intial iterations while is compatible with RAMLA. However, in contrast to RAMLA, and similar to COSEM, the new algorithm does not require any user-specified relaxation parameters.

For the future work, we will extend the proposed algorithm into a MAP case by including a prior.

Run COSEM few iterations and turn to HORSE later...

V. Acknowledgments

This work was supported by NIH NS32879, and by NSC 91-2320-B182-036 from NSC Taiwan.