SCALED HEAVY-BALL ACCELERATION OF THE RICHARDSON-LUCY ALGORITHM

Hongbin Wang, Paul Miller

The Centre for Secure Information Technologies (CSIT)
Queen's University of Belfast, Belfast, BT3 9DT, UK

ABSTRACT

The Richardson-Lucy algorithm is one of the most important algorithms in the image deconvolution area. However, one of its drawbacks is slow convergence. A very significant acceleration is obtained by the technique proposed by Biggs and Andrews (BA), which is implemented in the *deconvlucy* function of the *Image Processing* MATLAB toolbox. The BA method was developed heuristically with no proof of convergence. In this paper, we introduce the Heavy-Ball (H-B) method for Poisson data optimization and extend it to a scaled H-B method, which includes the BA method as a special case. The method has proof of the convergence rate of $O(k^{-2})$, where k is the number of iterations. We demonstrate the superior convergence performance of the scaled H-B method on both synthetic and real 3D images.

Index Terms— Deconvolution, Poisson noise, Richardson-Lucy algorithm, Acceleration, Heavy-ball method

1. INTRODUCTION

Restoration of 3D confocal microscopy images to remove blur and noise is an important preprocessing step for automated analysis and also for visualization by an end-user. One characteristic of 3D confocal microscopy imagery is that it is degraded mainly by Poisson noise under low light conditions. Perhaps the best well-known of restoration techniques for Poisson data is the Richardson-Lucy (RL) algorithm [1, 2]. However, the RL algorithm has drawbacks in terms of the quality of the restoration, and also the implementation time for 3D images. To address the slow convergence problem of the RL algorithm many acceleration methods have been proposed. Unfortunately, in many cases, the reduction in the number of iterations is achieved only at the expense of increased computational cost per iteration, resulting in an insignificant speed-up [3].

There are two exceptional pieces of work that have achieved significant acceleration for the RL algorithm. One of these methods, based on extrapolation of the iterative point, was proposed by Biggs and Andrews (BA) in [4]. A shortcoming of BA is that no convergence proof is available.

Another method, called scaled gradient projection, was proposed by Bonettini in [5]. It achieved similar performance with the BA method and has a convergence proof. However, it does not have a convergence rate proof.

In this work, we link the BA method to an optimization technique called the Heavy-Ball (H-B) method [6], and incorporate the idea of a scale matrix [5, 7] into it. The new technique, called scaled H-B method, includes the BA method as a special case. By using the theoretical proof that the negative Poisson log-likelihood has a Lipschitz-continuous gradient if the variable value is within a positive interval, and replacing the Lipschitz constant with a positive definite matrix, we prove that the scaled H-B method does not just converge, but that it has a faster convergence rate, of $O(k^{-2})$, than that of the original RL algorithm, of $O(k^{-1})$. We also empirically evaluate the scaled H-B method through experiments.

2. PROBLEM FORMULATION

Given an object x and its image y, which is acquired through an imaging system with a point-spread function (PSF) K and corrupted by a Poisson noise process P, then the image formation model is:

$$y = P\left(K \otimes x + b\right) \tag{1}$$

where $b \geq 0$ is a constant background and $\otimes$ denotes the convolution operation.

From a Bayesian viewpoint, the conditional probability of image y given object x is as follows:

$$p\left(y|x\right) = \prod_{i=1}^{N} \frac{\left(K \otimes x + b\right)^{y} e^{-\left(K \otimes x + b\right)_i}}{y!} \tag{2}$$

where N is the number of pixels in the image. Taking $C\left(x\right) = -\log p\left(y|x\right)$ as a cost function, and minimizing w.r.t x, gives the RL iterative equation

$$x_{k+1} = x_k \left(K * \frac{y}{K \otimes x_k} \right) \tag{3}$$

where $*$ is the correlation operator. Since the convergence of the RL algorithm is very slow, an acceleration method was proposed by Biggs and Andrews in [4]. During each iteration of the BA technique, a predicted point p_k, instead of x_k, is used to calculate x_{k+1} in Eqn (3), given by

$$p_k = x_k + \beta_k h_k \tag{4}$$

with $h_k = x_k - x_{k-1}$. β_k is the acceleration parameter, which can be calculated as follows:

$$\beta_k = \frac{\sum g_{k-1} g_{k-2}}{\sum g_{k-2} g_{k-2}} \tag{5}$$

where $g_k = x_{k+1} - p_k$ and $0 < \beta_k < 1$.

3. HEAVY-BALL METHOD

In this section, we firstly introduce the H-B method, and then we use two theoretical techniques to make all of the theoretical results of the H-B method available to us. Finally, we propose a generalized H-B, called scaled H-B, and derive the BA method as a special case.

3.1. The Heavy-Ball method for Poisson data

We begin by considering Eqn. (3) as a special case of the gradient projection method [3, 5]. Given the same cost function used for RL, the general iteration equation of the gradient projection method is as follows:

$$x_{k+1} = P_\Omega\left(x_k - \alpha_k \nabla C\left(x_k\right)\right) \tag{6}$$

where P_Ω denotes projection onto Ω, the feasible set of x and α_k is the kth step size. The RL algorithm can be interpreted as follows; P_Ω is the nonnegative constraint, the gradient of the cost function is $\nabla C(x_k) = 1 - (K * y / K \otimes x_k)$ and setting $\alpha_k = x_k$ gives Eqn (3).

The H-B method is a general acceleration method for gradient projection. It was proposed by Polyak [6], and the basic idea is to enhance the iterative updating by adding a momentum term from the previous step. The method can be separated into an extrapolation step

$$p_k = x_k + \beta_k \left(x_k - x_{k-1}\right) \tag{7}$$

And a gradient projection step

$$x_{k+1} = P_\Omega\left(p_k - \alpha_k \nabla C\left(p_k\right)\right) \tag{8}$$

Normally the gradient based method can only achieve a convergence rate of $O\left(k^{-1}\right)$ [8]. However, it can be proved that the H-B method achieves a best convergence rate of $O\left(k^{-2}\right)$ for the gradient based method [8] assuming:

1. The cost function $C\left(x\right)$ is convex and differentiable, and $x \in X$, where X is a closed convex set.
2. $C\left(x\right)$ has a Lipschitz continuous gradient: $\left\|\nabla C\left(x_1\right) - \nabla C\left(x_2\right)\right\| \le L\left\|x_1 - x_2\right\|, \forall x_1, x_2 \in X$, where L is Lipschitz constant.
3. $\alpha_k = 1/L$ and $\beta_k = \theta_k(1-\theta_{k-1})/\theta_{k-1}$, where the sequence $\left\{\theta_k\right\}$ satisfies $\theta_0 = \theta_1 \in (0,1]$, $\left(1-\theta_{k+1}\right)/\theta_{k+1}^2 \le 1/\theta_k^2$, and $\theta_k \le 2/\left(k+2\right)$, where k is the number of iterations. One possible choice is

$$\beta_k = \frac{k-1}{k+2}, \theta_k = \frac{2}{k+2} \tag{9}$$

Provided these assumptions are valid then we have:

Theorem 1 [8]: Let $\left\{x_k\right\}$ be a sequence generated by the H-B method, described by Eqns (7) and (8), and x^* be any optimal solution, then

$$C\left(x_k\right) - C\left(x^*\right) \le \frac{2L}{\left(k+1\right)^2}\left\|x_0 - x^*\right\|^2, \forall k \ge 1 \tag{10}$$

To prove this, our problem definition has to fulfill the above assumptions. Regarding the first assumption, the Poisson negative log-likelihood function is obviously convex and differentiable. By selecting appropriate values for α_k and β_k the 3rd assumption can be made to be met. Generally, however, the Poisson negative log-likelihood function does not have a Lipschitz continuous gradient because of the presence of the logarithm. Several previous works have tried to attack this problem, either through proposing new methods, such as the alternating direction optimization [9], or through transforming the Poisson distribution to a Gaussian [10].

Fortunately, we have noticed that, for image deconvolution the pixel value normally lies in a positive interval. So it is reasonable to enhance our nonnegative constraint to positive $x > a \ge 0$, where a is a constant. Therefore, we have the following lemma to satisfy the 2nd assumption.

Lemma 1: The negative Poisson log-likelihood has a Lipschitz-continuous gradient if the value of x lies within a positive interval.

Proof is omitted as a similar result can be found in [11].

3.2. Scaled Heavy-Ball and justification of BA method

To justify the BA method, we propose a slightly general form of H-B, called scaled H-B, which replaces the Lipschitz constant with a positive definite matrix to set the step size $\alpha_k = x_k$, thereby achieving better convergence. A similar idea has been explored in previous work [5, 7], however, we adapt it here to generalize the H-B method. The logic is that the Lipschitz gradient condition (2nd assumption) is required to guarantee the following inequality [8]

$$C\left(x_1\right) \le C\left(x_2\right) + \left\langle x_1 - x_2, \nabla C\left(x_2\right)\right\rangle + \frac{L}{2}\left\|x_1 - x_2\right\| \tag{11}$$

And that the inequality above is key to the proof of Theorem 1, which leads to setting the step size $\alpha_k = 1/L$. However the Lipschitz constant, L, is a loose estimate. If we can find a smaller Lipschitz constant, then, according to inequality (10), we will obtain faster convergence. This can be achieved through setting $\alpha_k = x_k$ and introducing a corresponding positive definite matrix $D = diag(x_k)$ to replace the Lipschitz constant. So inequality (11) generalizes to

$$C(x_1) \le C(x_2) + \langle x_1 - x_2, \nabla C(x_2) \rangle + \frac{1}{2}\|x_1 - x_2\|_D \quad (12)$$

In a manner similar to [7], we can replace inequality (10) in the proof of **Theorem 1** in [8], with inequality (12) to get the following theorem for the scaled H-B method.

Theorem 2: Let $\{x_k\}$ be a sequence generated by the scaled H-B method, and x^* be any optimal solution, then

$$C(x_k) - C(x^*) \le \frac{2}{(k+1)^2}\|x_0 - x^*\|_D^2, \forall k \ge 1 \quad (13)$$

Lemma 2: Scaled H-B converges faster than H-B.

Proof: ***Theorems 1*** *and* ***2*** *imply a convergence rate of* $O(k^{-2})$, *but* $\|x\|_D \le L\|x\|$. *So scaled H-B converges faster.*

We now provide a theoretical basis for the BA method, in that we can say it is a partially scaled H-B method. This is because it only replaces x_k with p_k inside the gradient as follows:

$$x_{k+1} = P_\Omega[x_k - \alpha_k \nabla C(p_k)] \quad (14)$$

Note that the BA method also has its special acceleration parameter defined by Eqn (5), however, it is not easy to prove that $\beta^{BA} \le \frac{k-1}{k+2}$. We can ensure the inequality by modifying β^{BA} as follows:

$$\beta^{BA} = \min\left\{\frac{\sum g_{k-1}g_{k-2}}{\sum g_{k-2}g_{k-2}}, \frac{k-1}{k+2}\right\} \quad (15)$$

Thus, we know that **Theorem 2** holds for the BA method.

4. EXPERIMENTS

In this section we demonstrate the performance of the scaled H-B method for different parameter choices on both synthetic and real 3D images.

4.1. Choosing step size and acceleration parameter

The two parameters in the scaled H-B method are step size α_k and acceleration parameter β_k. In the original H-B method [8], α_k is either set to $1/L$, which is seldom known in practice, or it is estimated using a backtracking strategy, which incurs extra computational cost. The advantage of scaled H-B is that it not only converges faster than the original version (**Lemma 2**), but, in practice, it is easily setup without additional cost by simply setting $\alpha_k = x_k$.

For the purposes of this experiment, we will evaluate three acceleration parameters; β_k^{BA} defined by Eqn (15), β_k^{H-B}, Eqn. (9) and β_k^{FISTA}, which is from the FISTA algorithm [12] and is defined as:

$$\beta_k^{FISTA} = \frac{t_k - 1}{t_{k+1}} \quad (16)$$

where $t_{k+1} = \left(1 + \sqrt{1 + 4t_k^2}\right)/2$. In fact, the FISTA algorithm can also be shown to be a special case of the scaled H-B method with $\alpha_k = 1/L$. For completeness, we will also evaluate two reference algorithms, RL and BA.

All algorithms are implemented in Matlab 7.12.0 and the BA method is the implementation in the *deconvlucy* function of the *Image Processing* MATLAB toolbox. The experiments are run on a Windows 7 64-bit machine with a processor Intel Core i7 2.80 Ghz and 6GB RAM.

4.2. Results

The synthetic dataset we use is called "hollow-bars" [13]. The 3D image, of size $256 \times 256 \times 128$, consists of six hollow bars arranged along a diagonal slice through a rectangular volume. As we already know the reference image x^*, because it is synthetic, the performance measure is defined as the KL divergence between x_k and x^*, denoted by $D_{KL}(x_k, x^*)$. To measure the speed-up factor, we let the RL algorithm run for 200 iterations and get a value for $D_{KL}(x_{200}, x^*)$. We then determine the number of iterations and time it takes for the other four algorithms to obtain a value smaller than $D_{KL}(x_{200}, x^*)$. The speed-up results are shown in Table 1 and the visual restoration results are presented in Fig 1.

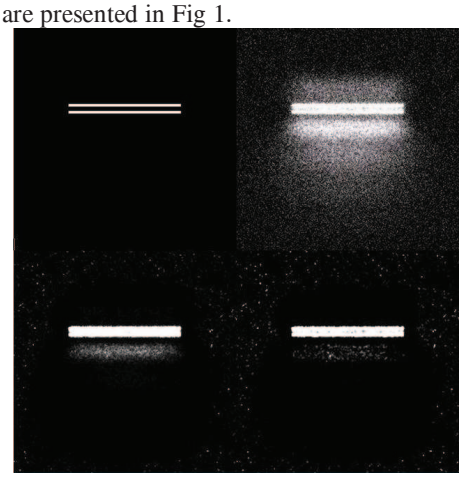


Fig 1. One stack view of "hollow-bars"(z=49). (Upper left) Original image; (Upper right) Blurred noisy image; (Lower left) Restoration by RL; (Lower right) Restoration by $H-B_\beta^{H-B}$.

Algorithm	Iteration	Time (sec)	Speed-up
RL	200	207.11	1.00
BA	29	38.22	5.42
$H-B_\beta^{BA}$	34	45.48	4.55
$H-B_\beta^{H-B}$	38	39.71	5.22
$H-B_\beta^{FISTA}$	38	39.91	5.19

Table 1. Speed-up of scaled H-B on synthetic data.

With similar visual restoration results from RL and scaled H-B in Fig 1, we observed a superior acceleration result from scaled H-B in Table 1. The acceleration parameter β_k^{BA} incurs extra computational cost to calculate, but reduces the number of iterations to 34. Both β_k^{H-B} and β_k^{FISTA} are trivial to calculate, but need slightly more iterations, 38. Overall, these three scaled H-B variants and the BA algorithm achieve similar speed-ups.

The real dataset used is the CY3 channel image of a C. Elegans embryo [13], which is of size 672x712x104. We run the same test procedure but the performance measure in this case is $C(x)$. The result is as follows:

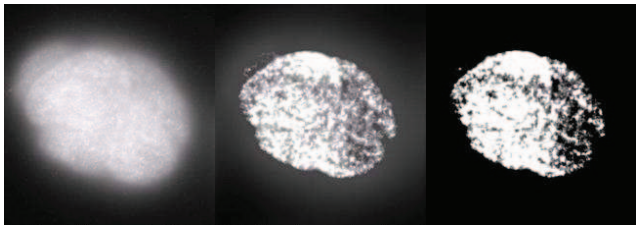


Fig 2. One stack view of real 3D image (z=80). (Left) Blurred noisy image; (Middle) Restoration by RL; (Right) Restoration by $H-B_\beta^{H-B}$

Algorithm	Iteration	Time (sec)	Speed-up
RL	200	1926.15	1.00
BA	0	Out of memory	0
$H-B_\beta^{BA}$	0	Out of memory	0
$H-B_\beta^{H-B}$	37	366.12	5.26
$H-B_\beta^{FISTA}$	37	367.95	5.23

Table 2. Speed-up of scaled H-B on real data.

We observed a similar result for the real image as for the "hollow-bars" image. In addition, we discovered that for larger images, the calculation of β_k^{BA} can cause memory problems.

5. CONCLUSIONS

The contribution of this paper is to propose the scaled H-B method to accelerate the RL algorithm. The method includes the BA method as a special case and has a convergence rate of $O(k^{-2})$. Experiments demonstrate its superior performance. In the future, we'd like to extend the analysis here to include a regularization term.

ACKNOWLEDGEMENTS

The authors would like to acknowledge funding from Invest NI grant ST285 and Andor Technology.

REFERENCES

[1] W. H. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. A, vol. 62, no. 1, pp. 55–59, 1972.

[2] L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron.J, vol.79, no.6, pp.745–754, 1974.

[3] M. Bertero, P. Boccacci, G. Desider↕ a, and G. Vicidomini, "Image deblurring with Poisson data: from cells to galaxies," Inverse Problems, 25(12):123006, 2009.

[4] D.S.C. Biggs and M.Andrews, "Acceleration of iterative image restoration algorithms," Appl. Opt.36, 1766–75, 1997.

[5] S.Bonettini, R.Zanella and L.Zanni, "A scaled gradient projection method for constrained image deblurring," Inverse Problems, 25, 015002, 2009.

[6] B.T.Polyak, "Introduction to Optimization, Optimization," Software. Inc., N.Y., 1987.

[7] Wangmeng Zuo and Zhouchen Lin, "A Generalized Accelerated Proximal Gradient Approach for Total-Variation-Based Image Restoration," IEEE Trans. Image Process., vol. 20, no.10, pp. 2748–2759, Oct. 2011.

[8] Dimitri P. Bertsekas, "Convex Optimization Theory, Supplementary Chapter 6 on Convex Optimization Algorithms," Athena Scientific, 2009.

[9] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," IEEE Trans. Image Process., vol. 19, no.12, pp.3133–3145, 2010.

[10] F.-X. Dup´e, J. Fadili, J.-L. Starck, "A proximal iteration for deconvolving Poisson images using sparse representations", IEEE Trans. on Image Proc., vol. 18, pp. 310–321, 2009.

[11] T.Zachary, R. F. M. Harmany and M. Rebecca , "SPIRAL out of Convexity: Sparsity-regularized Algorithms for Photon-limited Imaging," SPIE Electronic Imaging, 2010.

[12] A. Beck and M. Teboulle, "A fast iterative shrinkage thresholding algorithm for linear inverse problems," SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

[13] http://bigwww.epfl.ch/deconvolution/?p=