Correspondence

Simultaneous Perturbation Optimization for Efficient Image Restoration

The Gibbs sampler of Geman and Geman (G&G) (1984) is a standard technique for image restoration. We implemented a similar estimation technique based on a new optimization method (simultaneous perturbation stochastic approximation (SPSA)). Preliminary tests of our results show comparable performance to the Gibbs sampler but with faster running times. This study suggests that SPSA is a strong candidate for use in image restoration.

I. INTRODUCTION

Techniques of image processing and pattern recognition are in widespread use in defense, manufacturing, medical applications, and control [4]. Automated image processing algorithms generally are comprised of several steps such as image segmentation, object detection, and classification. We examine the process of image restoration, which refers to the task of recovering an image from a given degraded image (in pixel form), where some knowledge is available regarding the degradation process. We distinguish this from the process of image enhancement, which applies to various manipulations (such as modifying the contrast between the pixels of the image) that attempt to make the resulting image more pleasing to the human eye, but usually without any assumptions on the degradation process. Our objective is to investigate the applicability of an advanced optimization technique to the restoration process, and to compare the performance of our approach to that of a state-of-the-art restoration methodology.

II. IMAGE RESTORATION

In image restoration problems, the assumed degradation process usually involves a distorting function and random noise acting on the original image. Then it is appropriate to apply statistical

estimation techniques to try to arrive in some optimal way at the image that most likely generated the given degraded image. A popular technique for this application is Bayesian maximum a posteriori (MAP) estimation, which has received much attention in recent years [17 and the references therein]. We are concerned here mainly with the case where the given image is blurred by a convolution plus added Gaussian noise. This assumption is often made in restoration applications, and corresponds to expected modes of degradation. It is also possible (using arguments similar to those discussed below) to create methodologies based on criteria other than maximizing the Bayesian posterior distribution, such as maximum likelihood or conditional mean estimates. We focus on the MAP method because it provides an easy formalism for incorporating extra ("prior") information into the estimation process, and its computational demands are less than those of the conditional mean estimate.

In a landmark paper, Geman and Geman (G&G) [5] introduced the use of a sampling approach to image restoration, combined with stochastic relaxation (a type of "annealing") to produce the MAP image. The sampling methodology used by G&G, called the "Gibbs sampler," iteratively updates the pixel values, providing a recursion that converges to the MAP estimate of the original image. For ease of discussion, we restrict attention to the simple degradation model mentioned above, but the methodology also applies to more general models, e.g., where the noise is non-Gaussian or multiplicative and the distortion function is nonlinear (see [5, p. 723]).

The annealing method in G&G controls the shape of the Gibbs distribution (by decreasing a parameter called "temperature," analogous to the annealing process used in metalworking) that is used in the algorithm, making it more "peaked" as the algorithm proceeds in time. This contrasts with the method used in the popular simulated annealing algorithm, where "annealing" refers to changing a step-size parameter over time. The objective of numerical annealing is to promote convergence to a global, rather than a local, optimum solution, in G&G's case to the global maximum (mode) of the posterior distribution. Their algorithm has enjoyed success and is often cited (e.g., [17 and 7]).

III. A NEW APPROACH

Our approach to MAP estimation for image restoration is based on the same framework as that used by G&G, but replacing the Gibbs sampler with a relatively new optimization procedure called simultaneous perturbation stochastic approximation (SPSA), introduced by [10] and [11]. The implementations of the two methods are discussed in detail in Section IV, below. SPSA is an optimization technique based on stochastic approximation in the Kiefer-Wolfowitz setting. We chose to investigate the application of SPSA optimization methodology to the image restoration problem because it is easy to implement and has been shown to be efficient in other large-scale optimization problems. A related application of SPSA to a pattern recognition problem is discussed in [8], and an application to processing magnetospheric images is discussed in [3]. The efficiency of SPSA arises from the use of a special "simultaneous perturbation" gradient approximation that requires very few loss-function computations per iteration. The loss function used in the SPSA optimization for MAP image restoration is a quantity called the "energy function" that is closely related to the posterior probability function (see Section IV below). Minimizing the energy function is equivalent to maximizing the posterior.

SPSA is described in detail in [11] (see also www.jhuapl.edu/SPSA). For completeness, we provide a summary here. SPSA uses the standard stochastic approximation algorithm structure:

$$\omega_{k+1} = \omega_k - a_k \hat{g}_k(\omega_k)$$

where ω_k represents the current estimate of the MAP image as a vector of all the pixel intensities, $\{a_k\}$ is a "gain" sequence (described further below), and $\hat{g}_k(\omega)$ is an approximation to the gradient of the loss (or "energy") function $U^P(\omega)$ defined in Section IV. In standard implementations, $\{a_k\}$ is a sequence of positive numbers that decreases to zero, often chosen to be of the form $a_k = a/(k + A)^\alpha$, where a, A, and α are positive constants and $\alpha \leq 1$. The novel feature of SPSA is the so-called "simultaneous perturbation" gradient approximation, which is defined as

$$\hat{g}_k(\omega_k) = (2c_k \Delta_k)^{-1}[U^P(\omega_k + c_k \Delta_k) - U^P(\omega_k - c_k \Delta_k)]$$

(1)

where $\{c_k\}$ is a decreasing sequence of positive numbers, usually of form $c_k = c/k^\gamma$, with $c > 0$, and $0 \leq \gamma < \alpha$; Δ_k is a vector the same size as ω_k, say a p-vector; and the inverse of a vector is defined to be the vector of inverses ($\Delta_k = (\Delta_{k1}, \Delta_{k2}, \ldots, \Delta_{kp})^{\mathrm{T}} \Rightarrow \Delta_k^{-1} = (\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, \ldots, \Delta_{kp}^{-1})^{\mathrm{T}}$, where superscript T denotes transpose). The elements, Δ_{ki}, of Δ_k are chosen randomly according to the conditions in [11]. We often use the simple Bernoulli (± 1) distribution. (Uniformly or normally distributed perturbations are not allowed by the regularity conditions.)

Of course, the quantities $\omega_k \pm c_k \Delta_k$ will seldom form a vector of integer pixel levels. However, since the function $U^P(\omega)$ is differentiable for real vectors ω (using $V_C(\omega) = 0$ for the $\hat{g}_k(\omega_k)$ computation, with negligible loss of generality), this causes no problem for the optimization performed by SPSA. After SPSA develops ω_{k+1}, the progress of the energy function

is evaluated by rounding the elements of ω_{k+1} to the nearest integer contained in the allowable set $\{1, 2, \ldots, 11\}$. Other implementations of SPSA are possible for more general applications involving a discrete parameter (ω), e.g., [6].

Examination of definition (1) reveals that only two evaluations of the loss (energy) function are required in order to calculate this gradient approximation, in contrast to the "finite difference" gradient of elementary calculus, which would use $p + 1$ evaluations (or $2p$, for a two-sided approximation). The power of SPSA lies in the fact that, for optimization, this rough gradient approximation is nearly as informative as the standard gradient approximation. This surprising phenomenon has both theoretical underpinning [11], and empirical experience to support it. There is a large body of theory covering the convergence properties of various forms of SPSA [11, 2, 14 and the references therein], successful applications [12, 16] and helpful hints for implementation [13].

We mentioned above that the objective of G&G's annealing feature is convergence to the global optimum (MAP) solution. In fact, G&G proved a global convergence theorem for their methodology. However, the conditions for this convergence are often not met in practical applications (the starting temperature suggested by theory is so large that it would take impractical computation times to arrive at temperatures that are empirically known to be effective [5, p. 732]). SPSA also converges to a global optimum solution, under conditions developed in two theorems established in [9]. In addition, several numerical studies have demonstrated global convergence of SPSA [9, 1]. An intuitive explanation of this global convergence performance of SPSA is that the rough gradient approximation used in the algorithm introduces enough jitter into the early iterations to allow the search to escape local minima. Eventual convergence is enabled by the decreasing nature of the gain and $\{c_k\}$ sequences.

It is difficult to make analytical comparisons of the two approaches (Gibbs sampler and SPSA) for arriving at the MAP estimate, since their formal structures differ. As noted above, SPSA updates the entire image at each iteration, using a standard "stochastic approximation" recursion in tandem with its special gradient approximation. In contrast, the Gibbs sampler updates its estimate of the MAP image one pixel at a time while cycling repeatedly through all of the pixels. The update of a pixel is done by sampling from the set of pixel intensity levels according to probabilities defined as follows. Suppose that the sth pixel is to be updated. Let Λ represent the set of possible pixel intensity levels, ω_k the current Gibbs sampler estimate of the MAP image (at the kth update), and ω_k^x the image ω_k with the sth pixel replaced by $x \in \Lambda$. For every $x \in \Lambda$, probabilities $P_k(x)$ are calculated as follows:

$$P_k(x) = Z_s^{-1} \exp(-U^P(\omega_k^x)/T)$$

where $U^P(\bullet)$ is the energy function and T is the current value of the annealing temperature, both described in Section IV, and

$$Z_s = \sum_{y \in \Lambda} \exp(-U^P(\omega_k^y)/T).$$

The updated value of the sth pixel is generated by sampling $x \in \Lambda$ using the probabilities $P_k(x)$.

Some rough comparisons of the Gibbs sampler to SPSA can be made from our experience in the numerical example described below. G&G require several evaluations of the function $\exp(-U^P(\omega)/T)$ to update each pixel in a sweep through all the pixels, and the Gibbs sampler makes several sweeps. However, they are able to take advantage of a simplification in the computation (using only cliques containing neighbors of the pixel being updated) due to the local nature of the assumed Gibbs distributions [5, p. 730]. In our example, we found that the computation of the energy function by SPSA took about four times as long as G&G's simplified computation of the posterior. Also, one sweep by the G&G sampler took about 100 times as long to compute as one iteration of SPSA, but the G&G sweep made more progress towards the solution. While these results give a feel for the computational demands of the two algorithms, they give no obvious indication of the comparative performance, which is discussed further in Section IV below.

IV. NUMERICAL STUDY

The goal of this work is to demonstrate that SPSA should be considered as an optimizer for use in image restoration. We have already noted that SPSA's design and previous performance experience are strong indicators for SPSA's suitability for image processing. In addition, a theoretical comparison [15] of SPSA relative to a generic simulated annealing algorithm indicates that SPSA is likely to provide superior optimization performance. For a numerical demonstration, we chose G&G's implementation of the Gibbs sampler as a standard algorithm to compare with. Although G&G's original work is somewhat dated, their explanation of the algorithm is particularly clear. Our goal here is to show that SPSA can, in fact, outperform such a standard method using the same input image and the same basic assumptions.

We implemented the G&G Gibbs sampler restoration method along with an SPSA-based version of the same Bayesian MAP approach. The implementations of the two methods largely followed the description in G&G, using only the pixel intensity process (G&G also allow a treatment of lines in the

image, which we suppressed). For fairness in the comparisons, these implementations were as much alike as possible. In particular, the following elements were the same for our setups of both SPSA and the G&G Gibbs sampler:

1) The vector, say ω, of intensities (gray levels) of all of the pixels comprising the image, is modeled as a Markov random field, which is equivalent to assuming that the probability law for ω is a Gibbs distribution. For more details, see the discussion in [5, p. 722]. The definition of the Gibbs distribution [5, p. 725] is based on a system of neighborhoods, one for each pixel, and on cliques, which are sets of points, all of which are neighbors of each other [5, p. 724].

2) The neighborhood system consists of the eight nearest neighbors to each point, with the appropriate natural adjustments at the boundaries of the image; this is the neighborhood system shown in [5, Fig. 1(b)].

3) The cliques are determined by the neighborhoods, and are shown in [5, Figs. 1(d) and 1(e)].

4) The degraded image (to be restored) is obtained by convolving the original image with the shift-invariant blurring matrix, or point-spread function H defined in [5, eq. (2.3)], and adding Gaussian $N(\mu, \sigma^2)$ noise. The effect of the blurring matrix convolution is to replace each point of the original image by weighting the original intensity equally with the average of those of the eight nearest neighbors. We took $\mu = 0$, $\sigma = 1$.

5) The Markov random field and Gibbs distribution assumptions are as described in [5]. The upshot of this is that the posterior probability $P(\omega \mid g)$, of an image ω, where g is the vector of intensities of all of the pixels comprising the observed image (i.e., the given degraded image), is given by the Gibbs distribution [5, p. 728]:

$$P(\omega \mid g) = Z^{-1} \exp(-U^P(\omega)/T)$$

where T is an annealing temperature (discussed below), Z is a normalizing constant, and U^P, the "energy function," is given by [5, eq. (8.2)]:

$$U^P(\omega) = \sum_{C \in \mathbf{C}} V_C(\omega) + \|\vec{\mu} - (g - H(\omega))\|^2 / 2\sigma^2$$

where $H(\omega)$ is the image that results from applying the point-spread function to ω, $\vec{\mu}$ is a vector having every component equal to μ, $\| \bullet \|^2$ is the vector inner product (sum of squares of the components), $\mathbf{C}$ is the set of cliques, and V_C, the "potential" for any clique C is as described in [5, p. 732 under "Group 1"]:

$V_C(\omega) = 1/3$, if C is a clique of two elements, say $\{r, s\}$, and the r, s components of ω are equal,

$V_C(\omega) = -1/3$ in the same situation except that the r, s components of ω are different,

$V_C(\omega) = 0$, otherwise.

Regarding the temperature T we adopted a suggestion of [5, p. 732] of having T traverse from approximately $T = 4$ to $T = 0.5$ over 300 sweeps, by taking $T = 3/\log(1 + k)$, where k is the sweep number and goes from 1 to 300. We emphasize that the setups (input, image, Gibbs distribution, energy function, etc.) for both algorithms were as much alike, and as close to that specified by G&G, as possible.

For the implementation of SPSA, we used $T = 1$, along with the setup described above in Section III, with $a = 1.5$, $A = 20$, $\alpha = 0.602$, $c = 0.001$, and $\gamma = 0.101$. A discussion of the choice of parameters to use for SPSA is given in [13]. Another point about the implementation of the SPSA algorithm is that we used the following option, which often improves performance. The option is to prevent any update of the image that does not lower the cost function. Using this option requires an extra evaluation of the loss function at each iteration (in addition to the two evaluations that are required by the basic SPSA algorithm), in order to see whether the next image decreases the loss function value relative to the last update. If there is no decrease, the algorithm does not update the current image, and instead proceeds to the next iteration. Since the steps of the algorithm are based on a random process (used in approximating the gradient of the loss function), the next trial image will be different, and hence could decrease the loss function. While this idea of blocking updates that do not decrease the loss function is often a useful strategy, it can inhibit convergence to the true (global) minimum of the loss function by encouraging convergence to a local minimum. Using SPSA for convergence to the global minimum was discussed in Section III above.

To compare the two methods, we started with several 10×10 images (100 pixels) having 11 intensity levels (from zero to 10), and generated degraded images, blurred by the "point spread" convolution and additive noise described above. We applied both methodologies to restore (de-blur) the images. For each image, we tuned both the G&G and the SPSA algorithm by using a few test runs and observing how the loss function (i.e., the energy function, which both algorithms are trying to minimize) changed with various parameter settings. A comparison of typical final results is given in Fig. 1, which shows the original image (which, of course, is unknown to the algorithms), the degraded image (which is input to the algorithms), the G&G and SPSA solutions for the MAP image, and the loss function history over the span of the recursions. We found the SPSA approach to be competitive with the established method, both in finding an image that minimized the energy function (hence maximized the posterior conditional probability) and in the quality of the final image (measured subjectively by eye and also by summing the absolute differences of pixel

Original Image	G&G MAP	G&G Loss Fn.

300 SWEEPS

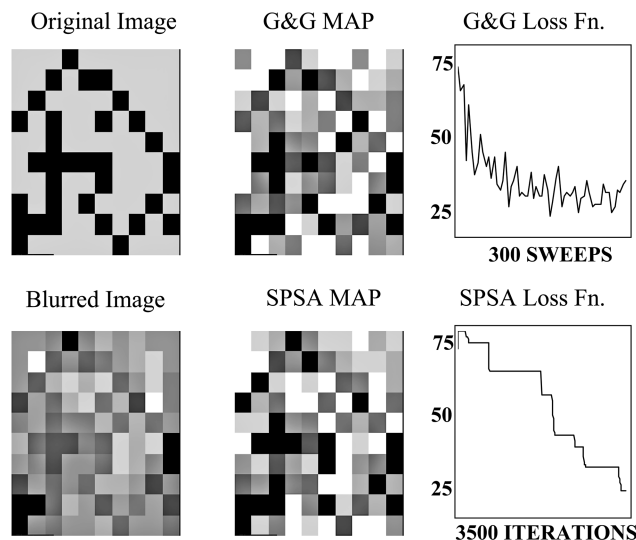| Blurred Image	SPSA MAP	SPSA Loss Fn.

3500 ITERATIONS

Fig. 1. Images and loss function values (same loss function for both; G&G 300 sweeps took eight times as long as SPSA 3500 iterations).

intensities between the original and the MAP images). Of course, since MAP is our criterion, the only true measure of the quality of the solution (final image) is the level of the final energy function. But the visual and pixel-difference comparisons provide further evidence that the two algorithms are arriving at very similar solutions.

The SPSA algorithm achieved these results in much faster execution times on several tests involving these relatively small images. Typical timing results saw 300 sweeps of the Gibbs sampler running about eight times as long as 3500 iterations of SPSA (we chose the running times so as to match the loss function performance of the two algorithms) on a Pentium-class computer. We tried some longer runs, but they did not improve the loss function perceptibly. These timing results should be interpreted with caution. The Gibbs sampler is an established and successful algorithm, and, although we implemented it carefully, results are likely to be implementation-dependent. We wrote both algorithms in Matlab, and ran them without using a compiler. While these results are preliminary, they indicate that the SPSA approach has the potential to improve on current technology.

V. SUMMARY

SPSA is an efficient methodology designed to handle complex optimization problems such as image restoration. We compared SPSA to the well-established Gibbs sampler of Geman & Geman. We found SPSA easier to implement than the Gibbs sampler since the SPSA algorithm has a simpler form. Our numerical comparisons showed that SPSA can provide comparable MAP solutions for image

restoration in much faster running times than the Gibbs sampler. Our experience with other applications indicates that the SPSA computational advantage may increase with larger images. This is because the number of loss function evaluations per iteration for SPSA does not increase with increasing image size and the required increase in number of iterations tends to be moderate. Further, a "second order" version of SPSA [14], not implemented in the preliminary study described here) has shown the ability to provide potentially dramatic further improvements in convergence without requiring information beyond that of the basic SPSA used here. SPSA's large body of supporting theory, its proven performance in other complex optimization applications, and its competitive performance with the established G&G algorithm provide strong indications of the suitability of SPSA for image processing.

JOHN L. MARYAK
JAMES C. SPALL
The Johns Hopkins University 24-w247
Applied Physics Laboratory
11100 Johns Hopkins Rd.
Laurel, MD 20723-6099
E-mail: (john.maryak@jhuapl.edu)
(james.spall@jhuapl.edu)

REFERENCES

[1] Chin, D. C.
A more efficient global optimization algorithm based on Styblinski and Tang.
Neural Networks, **7** (1994), 573–574.

[2] Chin, D. C.
Comparative study of stochastic algorithms for system optimization based on gradient approximations.
IEEE Transactions on Systems, Man and Cybernetics, Part B, **27** (1997), 244–249.

[3] Chin, D. C.
The simultaneous perturbation method for processing magnetospheric images.
Optical Engineering, **38** (1999), 606–611.

[4] Chou, K. C., Willsky, A. S., and Benveniste, A.
Multiscale recursive estimation, data fusion and regularization.
IEEE Transactions on Automatic Control, **39** (1994), 464–478.

[5] Geman, S., and Geman, D.
Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-6** (1984), 721–741.

[6] Gerencser, L., Hill, S. D., and Vago, Z.
Discrete optimization via SPSA.
In *Proceedings of the American Control Conference*, Arlington, VA, June 25–27, 2001, 1503–1504.

[7] Hurn, M., and Jennison, C.
An extension of Geman and Reynolds' approach to constrained restoration and the recovery of discontinuities.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **18** (1996), 657–662.

[8]   Maeda, Y., Hirano, H., and Kanata, Y.
       A learning rule of neural networks via simultaneous
       perturbation and its hardware implementation.
       Neural Networks, **8** (1995), 251–259.

[9]   Maryak, J. L., and Chin, D. C.
       Global random optimization by simultaneous perturbation
       stochastic approximation.
       In *Proceedings of the American Control Conference*,
       Arlington, VA, June 25–27, 2001, 756–762.

[10]   Spall, J. C.
       A stochastic approximation algorithm for
       large-dimensional systems in the Kiefer-Wolfowitz
       setting.
       In *Proceedings of the IEEE Conference on Decision and
       Control*, 1988, 1544–1548.

[11]   Spall, J. C.
       Multivariate stochastic approximation using a
       simultaneous perturbation gradient approximation.
       IEEE Transactions on Automatic Control, **37** (1992),
       332–341.

[12]   Spall, J. C.
       An overview of the simultaneous perturbation method for
       efficient optimization.
       Johns Hopkins APL Technical Digest, **19** (1998), 482–492,
       (http://techdigest.jhuapl.edu/td/td1904/spall.pdf).

[13]   Spall, J. C.
       Implementation of the simultaneous perturbation
       algorithm for stochastic optimization.
       IEEE Transactions on Aerospace and Electronic Systems,
       34 (1998), 817–823.

[14]   Spall, J. C.
       Adaptive stochastic approximation by the simultaneous
       perturbation method.
       IEEE Transactions on Automatic Control, **45** (2000),
       1839–1853.

[15]   Spall, J. C., Hill, S. D., and Stark, D. R.
       Some theoretical comparisons of stochastic optimization
       approaches.
       In *Proceedings of the American Control Conference*,
       Chicago, June 28–30, 2000, 1904–1908.

[16]   Whitney, J. E., Duncan, K., Richardson, M., and
       Bankman, I.
       Parameter estimation in a highly nonlinear model using
       simultaneous perturbation stochastic approximation.
       Communications is Statistics—Theory and Methods, **29**
       2000, 1247–1256.

[17]   Winkler, G.
       *Image Analysis, Random Fields, and Dynamic Monte Carlo
       Methods*.
       Berlin: Springer, 1995.