LEAST-SQUARE HALFTONING VIA HUMAN VISION SYSTEM AND MARKOV GRADIENT DESCENT (LS-MGD): ALGORITHM AND ANALYSIS

JIANHONG (JACKIE) SHEN *

Abstract.
Halftoning is the core algorithm governing most digital printing or imaging devices, by which images of continuous tones are converted to ensembles of discrete or quantum dots. It is through the human vision system (HVS) that such fields of quantum dots can be perceived almost identical to the original continuous images. In the current work, we propose a least-square based halftoning model with a substantial contribution from the HVS model, and design a robust computational algorithm based on Markov random walks. Furthermore, we discuss and quantify the important role of spatial smoothing by the HVS, and rigorously prove the gradient-descent property of the Markov-stochastic algorithm. Computational results on typical test images further confirm the performance of the new approach. The proposed algorithm and its mathematical analysis are generically applicable to similar discrete or nonlinear programming problems.

Key words. Halftoning, human vision system (HVS), spatial smoothing, entropy, least square, random fields, stochastic gradient descent, Markov random walks, Bernoulli flipping, stochastic convergence, blue noise

AMS subject classifications. 94A08, 62M40, 68U10

1. Introduction. Halftoning is the key algorithm built into most binary (with the black ink) or quintic (with the CMYK inks) printers, as well as pixel-based digital imaging devices such as CCD arrays based on Bayer tricolor filtering. To halftone an image u with *continuous* tones (or *contones*) is to find a suitable binary or quintic image b which can be perceived by the human vision system (HVS) almost identical to the original contone image u. The current work proposes a novel halftoning method for gray-scale images based on the least-square principle and Markov random walks, and also develops its mathematical analysis.

For real printers, halftoning must take into account the overlapping physical effects of adjacent ink droplets (i.e., the *printer* model [2, 50]). On the other hand, the growing interest in *pure* halftoning algorithms, i.e., the mathematical and computational study of contone-halftone conversions without considering the hardware issues, has been nourished by several novel applications in contemporary information technologies, including for example, digital watermarking, information hiding, and the sigma-delta model for Analog/Digital (A/D) conversion [1, 14, 15, 20, 25]. The current work also focuses only on the algorithmic development and analysis, and does not address any practical issues concerning hardware implementations.

The literature on halftoning is wealthy and highly diversified. There are mainly three leading classes of halftoning techniques, thanks to the collective efforts of many researchers:

[1] dot dithering or screening based on direct thresholding [3, 27, 49], one of the oldest halftoning approaches with the lowest computational complexities;
[2] error diffusion initiated by the classic work of Floyd and Steinberg [18] in 1976, and many of its improved versions [16, 17, 26, 48, 51, 52]; and
[3] dot diffusion based on domain tiling and class-matrix guided ranking [26, 30].

On top of these three major categories, there also exist several other techniques including for example, look-up tables, least-square minimization, Markov random fields, direct binary search (DBS), fuzzy algorithms, as well as the most recent model proposed by the author based on Perona-Malik's error diffusion and stochastic flipping [2, 21, 27, 32, 34, 39, 41, 47]. The references enumerated above are almost representative but by no means complete. We further refer the reader to the vast bibliographies therein.

Within the mathematics community, fresh interest in halftoning has been primarily driven by the innate quest for more rigorous analysis of halftoning algorithms. The main challenge of halftoning analysis is characteristically distinct from perhaps any other known class of image processing tasks, e.g., denoising, deblurring, dejittering, inpainting, or segmentation [5, 6, 7, 8, 22, 31, 35, 36, 37, 38, 40, 42]. By definition, halftoning lies on the gray buffer between the continuum and discrete realms. Mathematically, there have been no well developed tools for this frontier area of research, and analytical results have to be derived from multiple tools including probability theory, number theory, functional and harmonic analysis, linear

*This work has been partially supported by the NSF (USA) under grant number DMS-0604510. Address: Lotus Hill Institute for Computer Vision and Information Science, E'Zhou, Hu'Bei 436000, China; and Barclays Capital, Wall Street, New York, NY 10066, USA (jackieneoshen@gmail.com).

programming, approximation theory, and stochastic analysis. The recent wave of mathematical interest has been mainly stimulated by the remarkable works of Daubechies, DeVore, Güntürk, and their colleagues [13, 14, 15, 24, 25], especially on the coarse quantization and A/D conversion of analog signals.

The present work attempts to make the following contributions (throughout the six sections).

[1] We propose a novel halftoning algorithm based on the principle of least-square optimization and a stochastic marching strategy called Markov Gradient Descent (LS-MGD halftoning) (Sections 2, 3, 4). The principle of least-square optimization for halftoning has also been explored in some earlier works (see, e.g., Neuhoff, Pappas, and Seshadri [32] for one-dimensional halftoning). The principal novel contribution here is the MGD marching scheme, which amounts to a Markov random walk within the ensemble of all halftone images.

[2] We quantify and analyze the significant role of the Human Vision System (HVS) for halftoning design. HVS modeling has been an important component in the halftoning literature [30, 47, 53] (also see Section 2). However, it has never been made clear in the vast literature as to which property of the HVS significantly influences halftoning design and how. In the current work, we explicitly reveal the significance of the *spatial smoothing* property of an HVS model (Section 5).

[3] We develop the analytical results for the proposed LS-MGD algorithm (Section 6). The main theoretical results are written into the major theorem in Section 6, which is expressed in two different versions: one is semi-heuristic but much easier to comprehend, while the other is rigorous but more involved (Theorem 6.2). Such analytical results are rare in the halftoning literature, and also shed important light on the robust computational performance in Section 7.

As in our earlier work on Perona-Malik's error diffusion and stochastic halftoning [39, 41], the LS-MGD algorithm proposed herein is also *stochastic* since the iterative marching amounts to a Markov random walk within the ensemble of all halftone images. As a result, the final halftone outputs are always some *random* fields in stochastic equilibrium, instead of *deterministic* binary fields produced by most deterministic algorithms in existence.

Finally, in this relatively new area of halftoning or 2-D Analog/Digital conversion, the current work by no means has been perfected. Many intriguing aspects in terms of algorithm design or analysis still await to be developed. Together with all the remarkable works aforementioned, we hope that an interesting class of mathematical problems with important digital applications have been faithfully formulated and stimulated for the mathematics community.

2. Modeling the Human Vision System (HVS). We first briefly discuss one of the most important ingredients of halftoning design – the modeling of the HVS. More detailed analysis is developed further in Section 5.

Let u be a contone image defined on a uniform Cartesian lattice Ω. After scaling normalization, one could assume that $\Omega = [1:n] \times [1:m]$. In this paper, occasionally we also consider the infinite lattice $\mathbb{Z}^2$ to facilitate rigorous mathematical formulation or analysis.

Let the HVS be modelled by a linear operator $K : v \to K[v]$, for any image v on Ω whose gray-scale range is $[0, 1]$ (whether being contone or halftone). We shall call $K[v]$ the *perceived* image of v. Throughout the present work, the HSV K shall always be assumed *weakly lowpass* in the sense of $K[1] \equiv 1$, which is also called the direct-current (DC) condition in signal processing [29, 33, 46].

Empirically, it appears rational to assume that human visual perception is *independent* of, or at least remains *stable* with respect to eyeball motion or head turning, which implies that the HVS operator K should be *shift invariant* when the domain is ideally $\mathbb{Z}^2$. As well known in digital signal processing [33], such an HVS must be specified by a point spread function (PSF) $k_{i,j}$ and the convolution operator:

$$K[v]_{i,j} = (k * v)_{i,j} = \sum_{(s,t)\in\mathbb{Z}^2} k_{i-s,j-t} v_{s,t}.$$

For convenience, we shall use $\alpha, \beta, \ldots$ to denote the pixels. For instance, let $\alpha = (i, j)$ and $\beta = (s, t) \in \mathbb{Z}^2$. Then one can simply write

$$(k * v)_\alpha = \sum_\beta k_{\alpha-\beta} v_\beta.$$

The weak lowpass condition requires that $\sum_\alpha k_\alpha = 1$. One may also impose the *positivity* condition: $k_\alpha \geq 0$ for any pixel α, in order to

be consistent with the range assumption $u \in [0,1]$, so that $v = k * u$ also falls into the same canonical range of $[0,1]$.

Then the weak lowpass condition can be concisely expressed via the l^1-norm:

$$\|k\|_1 = \sum_\alpha k_\alpha = 1. \tag{2.1}$$

Furthermore, one can assume that the PSF k is symmetric: $k_{-\alpha} = k_\alpha$, which implies that the HSV K is a symmetric linear operator in $l^2(\Omega)$:

$$\langle K[u],\ v\rangle = \langle u,\ K[v]\rangle, \qquad \text{for any } u, v \in l^2(\Omega).$$

Hereafter, $\langle \cdot,\ \cdot\rangle$ always denotes the canonical inner product in $l^2(\Omega)$. Classical Gaussian and Cauchy kernels k^{g} and k^{c}, for instance, are indeed both positive and symmetric [19]:

$$k^{\mathrm{g}}_\alpha = \frac{1}{Z_b} e^{-b|\alpha|^2}, \qquad \text{and} \qquad k^{\mathrm{c}}_\alpha = \frac{1}{Z_c}\frac{1}{(c|\alpha|^2+1)^{3/2}}, \quad \alpha \in \mathbb{Z}^2, \tag{2.2}$$

where $b, c > 0$ are two controlling parameters, and Z_b and Z_c are the associated normalization constants to meet the weak lowpass condition (or, the so-called *partition functions* in statistical mechanics [10, 7, 23]).

3. Halftoning Error Field and Least Square Formulation. Let $|\Omega| = N = n \times m$ denote the number of pixels on the image domain. A halftone image b is a binary map or field:

$$b : \Omega \to \{0,1\}; \qquad \alpha \to b_\alpha \in \{0,1\},$$

and the ensemble of all halftone images can be represented by $\{0,1\}^\Omega$. As a result, the total number of halftone images is $2^{|\Omega|}$, which is astronomical for any domain of even a moderate size, e.g., 256 by 256.

Given a contone image v and a halftone image b, we define the *perceived* error field e to be:

$$e = e(b \mid v) = K[v] - K[b] = u - K[b], \qquad \text{with } u = K[v],$$

under a given HVS K. Once both v and K are given, the perceived image $u = K[v]$ is fixed and can be equally treated as a given input. Thus in what follows, we shall always work with the u-variable instead of the original image v. In particular, $e(b \mid v)$ shall also be rewritten as $e(b \mid u)$, and the same rule applies to any other quantity $f(X \mid v)$ depending on some X *given* the original image v. The goal of least-square halftoning is to find the optimal halftone image which minimizes the total squared errors:

$$D(b \mid u) = D_K(b \mid u) = \sum_{\alpha\in\Omega} (u_\alpha - K[b]_\alpha)^2 = \|e(b \mid u)\|_2^2. \tag{3.1}$$

The least square formulation (3.1) reminds one of the deblurring problem in image processing [7]. For deblurring, u represents the blurred image, and b the original clear (contone) image. When both u and b are contone images, the above least-square formulation usually has to incorporate the information of the *hypothesis space* $\mathcal{B}$ (or the model space) [12, 43] for b and bears the form of

$$\min_{b\in\mathcal{B}} D(b \mid u) + t\|b\|_{\mathcal{B}},$$

where $t > 0$ is a regularity weight, and $\|\cdot\|_{\mathcal{B}}$ denotes the norm or semi-norm of $\mathcal{B}$. We refer the reader to our most recent work [9] for a comprehensive review on model-based variational deblurring.

We emphasize that the halftoning problem could be treated as a deblurring problem in a unconventional sense. The perceived contone image u indeed results from the blurring effect of the human vision system, which is the foundation of the entire halftoning machinery. To any artificial vision system whose resolution

is as fine as the grain size of the image domain, a binary halftone image b would remain a turbulent field of quantum dots, and thus is useless for faithfully representing continuous image patterns.

For halftoning, the model space is $\mathcal{B} = \{0,1\}^\Omega$, which is discrete and does not have any obvious regularity measure $\|b\|_{\mathcal{B}}$. One natural metric is motivated by the celebrated Ising model for ferromagnets in statistical mechanics [10]. Define the spin transform $\{0,1\} \to \{-1,1\}$: $s_\alpha = 2b_\alpha - 1$. Then the Ising regularity is given by:

$$E_{\text{Ising}}(b) = -J \sum_{\alpha\sim\beta} s_\alpha s_\beta - H \sum_\gamma s_\gamma,$$

where J and H are two positive parameters (representing internal and external field strengths), and $\alpha \sim \beta$ means that α and β are neighbors (under a designated graph structure). However, this regulatory "energy" encourages the formation of chunky clusters, or *red* noises, which are undesirable for halftoning [48, 49]. Halftoning goes after the very opposite, i.e., neighboring tones should differ from each other whenever it is feasible.

In the current work, we focus on the design and analysis of a novel algorithm for solving the least square model (3.1) directly and stochastically, without the introduction of any regularity terms.

4. Halftoning via Least Squares and Markov Gradient Descent.

4.1. Motivation from the Classical Gradient Descent Scheme. We now study the halftoning model based on the least-square principle:

$$\min_{b\in\{0,1\}^\Omega} D(b \mid u) = \min_{b\in\{0,1\}^\Omega} \|u - K[b]\|_2^2, \tag{4.1}$$

for any given contone image u in the canonical range of $[0,1]$.

If b *were* also a contone image and allowed a continuum deformation: $b \to b + \delta b$ with $\|\delta b\|_2 \ll 1$, the first variation would then be given by

$$\delta D(b \mid u)\,(\simeq D(b+\delta \mid u) - D(b \mid u)) = 2\langle u - K[b],\ -K[\delta b]\rangle = -2\langle K^T[e],\ \delta b\rangle,$$

where $e = u - K[b]$ is the perceived error field, and K^T the adjoint or transpose of the linear HSV K. As a result, the classical gradient descent method would be given by

$$\frac{\partial b(t)}{\partial t} = K^T[e], \qquad \frac{\partial b_\alpha(t)}{\partial t} = K^T[u - K[b(t)]]_\alpha, \qquad t > 0, \qquad \alpha \in \Omega, \tag{4.2}$$

where t is the artificial marching time, and $b(t) = (b_\alpha(t) \mid \alpha \in \Omega)$ the approximation at time t. With an infinitesimal numerical marching step $\tau = \Delta t$, $b^n = b(n\tau)$, and $\Delta b^n = b^{n+1} - b^n$, one has

$$b^{n+1} = b^n + \Delta b^n = b^n + \tau K^T[e^n], \qquad n = 0, 1, \cdots, \tag{4.3}$$

started with some suitable initial guess. As well known in linear algebra [44, 45], in the limit the time marching solves the *normal equation* $K^T[e] = 0$.

Except for extremely lucky scenarios, however, this marching scheme dooms to fail for halftoning due to the discreteness of the admissible space $\{0,1\}^\Omega$. For halftoning, only quantum leaps are allowed at each pixel, i.e., either keeping the current value b_α^n or flipping it to the opposite. Continuous and infinitesimal marching as in (4.2) or (4.3) is thus prohibited.

4.2. Randomized Gradient Descent. Instead of the *deterministic* marching scheme (4.3), we propose to explore it in the spirit of *stochastic* evolution. Let b^n be the current halftone approximation of a given contone image u. Then at each pixel $\alpha \in \Omega$, the flipping amount $\Delta b_\alpha^n = b_\alpha^{n+1} - b_\alpha^n$ must satisfy

$$\Delta b_\alpha^n \in \{0 - b_\alpha^n, 1 - b_\alpha^n\}, \quad \text{in order to make sure that} \quad b_\alpha^{n+1} \in \{0,1\}.$$

This implies that

$$\text{either} \quad \Delta b_\alpha^n \in \{0,1\}, \quad \text{when } b_\alpha^n = 0; \quad \text{or} \quad \Delta b_\alpha^n \in \{0,-1\} \quad \text{when } b_\alpha^n = 1. \tag{4.4}$$

As a result, one can treat Δb_α^n as a binary random variable, and either Δb_α^n or $-\Delta b_\alpha^n$ must be Bernoulli. Then the classical marching formula $\Delta b_\alpha^n = \tau K^T[e^n]_\alpha$ can be understood stochastically as:

$$\mathrm{E}[\Delta b_\alpha^n] = \tau K^T[e^n]_\alpha, \qquad \alpha \in \Omega, \tag{4.5}$$

where $\mathrm{E}[X]$ denotes the expectation operator for a random variable X.

Recall that for a Bernoulli random variable $X \sim \mathrm{Bernoulli}(p)$, suppose

$$p = \mathrm{Prob}(X = 1), \quad \text{and} \quad q = 1 - p = \mathrm{Prob}(X = 0).$$

Then the mean and variance are given by

$$\mathrm{E}[X] = p, \quad \text{and} \quad \mathrm{var}(X) = pq = p(1-p). \tag{4.6}$$

As a result, (4.5) implies that

$$\begin{aligned} \mathrm{Prob}(\Delta b_\alpha^n = 1) &= \tau K^T[e^n]_\alpha, \qquad \text{if } K^T[e^n]_\alpha \geq 0; \qquad \text{and,} \\ \mathrm{Prob}(\Delta b_\alpha^n = -1) &= -\mathrm{E}[\Delta b_\alpha^n] = -\tau K^T[e^n]_\alpha, \qquad \text{if } K^T[e^n]_\alpha < 0. \end{aligned} \tag{4.7}$$

In order for these stochastic flipping rules to be meaningful in probability, one has to make sure that $p = \left|\tau K^T[e^n]_\alpha\right| \leq 1$, which is investigated by the next theorem. Immediately following this theorem, we shall also study the overall realizability of the above analysis.

THEOREM 4.1 (Permissible Range of Stochastic Marching Size τ). *Suppose the* HVS K *is symmetric and positive, i.e., for any* $v = (v_\alpha \mid \alpha \in \Omega)$ *with* $v_\alpha \geq 0$*, one always has* $K[v]_\beta \geq 0$ *for any* $\beta \in \Omega$*. Then for any given canonical contone image* u *and halftone image* b*, one must have* $\left|K^T[e]_\alpha\right| \leq 1$ *for any* $\alpha \in \Omega$*. In particular, the probability identities in* (4.7) *are realizable for any marching step size* $\tau \in (0, 1]$*. Finally, if* K *is shift invariant, the symmetry condition can be removed.*

Proof. Let χ denote the white image on Ω so that $\chi_\alpha \equiv 1$ for any $\alpha \in \Omega$ (in order to avoid the confusion between 1 as a scalar and 1 as an image). Then

$$\|v\|_\infty \chi_\alpha \pm v_\alpha \geq 0, \qquad \alpha \in \Omega.$$

By applying the positivity condition (which implies *monotonicity*), one has

$$-\|v\|_\infty K[\chi]_\alpha \leq K[v]_\alpha \leq \|v\|_\infty K[\chi]_\alpha, \qquad \alpha \in \Omega.$$

By the weak lowpass condition discussed earlier, $K[\chi] \equiv \chi \equiv 1$. Therefore,

$$\|K[v]\|_\infty \leq \|v\|_\infty. \tag{4.8}$$

For any contone image $u \in [0, 1]$ and halftone image $b \in \{0, 1\}$, one has as a result,

$$0 \leq K[b]_\alpha \leq \|b\|_\infty \leq 1, \qquad \alpha \in \Omega,$$

where the leftmost inequality is also due to the positivity condition. Consequently,

$$e_\alpha = u_\alpha - K[b]_\alpha \in [-1, 1], \qquad \alpha \in \Omega.$$

Since K is symmetric, one has

$$\|K^T[e]\|_\infty = \|K[e]\|_\infty \leq \|e\|_\infty \leq 1.$$

This completes the proof for a general HSV K that is symmetric and positive.

Finally, if K is shift invariant with PSF k, then the symmetry condition (which is equivalent to $k_{-\alpha} \equiv k_\alpha$) can be removed since the positivity of k immediately implies the positivity of $k_{-\bullet}$, which is the PSF of K^T. Then the last equation line still holds due to $\|K^T[e]\|_\infty \leq \|e\|_\infty$. □

The above theorem guarantees the stochastic realizability of the randomized gradient descent:

$$\mathrm{E}[\Delta b_\alpha^n] = \tau K^T[e^n]_\alpha.$$

However, not all samples from such random distributions are permissible due to the constraint

$$b_\alpha^{n+1} = b_\alpha^n + \Delta b_\alpha^n \in \{0, 1\}, \qquad \alpha \in \Omega.$$

For example, suppose at a pixel α, one has $b_\alpha^n = 0$ while $K^T[e]_\alpha < 0$. Then according to (4.7), Δb_α^n must be a Bernoulli random variable supported on $\{0, -1\}$. This would imply that with a positive chance the new halftone value $b_\alpha^{n+1} = b_\alpha^n + \Delta b_\alpha^n$ could be $0 + (-1) = -1$, which is prohibited. Similarly, the scenario of $b_\alpha^n = 1$ and $K^T[e]_\alpha > 0$ should also be precluded. This inspires the following notion of *compatibility*.

Definition 4.2 (Compatibility for Stochastic Gradient Descent). *A halftone image b and its error field $e = u - K[b]$ are said to be compatible at a pixel α if*

$$(2b_\alpha - 1) \cdot K^T[e]_\alpha \leq 0. \tag{4.9}$$

For example, if K is the identity operator, then b and e are compatible at *any* pixel. (One can further show that in this case our algorithm produces an image-driven random field of nonhomogeneous Bernoulli white noise.)

Proposition 4.3 (Bernoulli Formulae for Stochastic Gradient Descent). *Suppose b^n and e^n are compatible at a pixel α. Then the stochastic gradient descent* (4.7) *leads to a Bernoulli distribution for b_α^{n+1} with*

$$p_\alpha = \mathrm{Prob}(b_\alpha^{n+1} = 1) = b_\alpha^n + \tau K^T[e]_\alpha, \tag{4.10}$$

provided that K is positive and symmetric and $\tau \in (0, 1]$, as established in the preceding theorem.

Notice that due to the compatibility condition and Theorem 4.1, p_α defined in (4.10) indeed falls into the probability range $[0, 1]$.

Proof. Notice that $b_\alpha^{n+1} = 1$ if and only if $\Delta b_\alpha^n = 1 - b_\alpha^n$. Let p_α denotes $\mathrm{Prob}(b_\alpha^{n+1} = 1)$. Then

$$\mathrm{Prob}(\Delta b_\alpha^n = 1 - b_\alpha^n) = p_\alpha, \quad \text{and} \quad \mathrm{Prob}(\Delta b_\alpha^n = 0 - b_\alpha^n) = 1 - p_\alpha.$$

Therefore,

$$\mathrm{E}[\Delta b_\alpha^n] = p_\alpha(1 - b_\alpha^n) + (1 - p_\alpha)(0 - b_\alpha^n) = p_\alpha - b_\alpha^n.$$

By the stochastic gradient descent formula (4.5), one has

$$p_\alpha - b_\alpha^n = \tau K^T[e^n]_\alpha,$$

which is precisely (4.10). □

4.3. The Algorithm and Its Markov Interpretation. To summarize the above developments, we propose the new algorithm in the form of a pseudo code.

The Algorithm of Least-Square Based Markov Gradient Descent (LS-MGD):

(i) input the contone image u and marching step size $\tau \in (0, 1]$;
(ii) initialize the halftone b^0, e.g., $b_\alpha^0 \sim \mathrm{Bernoulli}(u_\alpha)$ at any pixel α;
(iii) **for** $n = 0, 1, \cdots$
  (iii-1) compute the perceived error field $e^n = u - K[b^n]$;
  (iii-2) compute the probability field $p = b^n + \tau K^T[e^n]$, and copy b^{n+1} from b^n;
  (iii-3) for those pixels α's with $p_\alpha \in [0, 1]$, flip $b_\alpha^{n+1} \in \{0, 1\}$ according to $\mathrm{Bernoulli}(p_\alpha)$.
  end

End of the Algorithm.

We make the following important remarks about LS-MGD halftoning.

[1] By Proposition 4.3 and the algorithm lines (iii-2) and (iii-3), b^{n+1} and b^n can only *potentially* differ on compatible pixels.

[2] Approaching stochastic convergence (as discussed in detail in the coming sections), the perceived error fields e^n's have very small magnitudes. As a result, the chance for b_α^{n+1} to differ from b_α^n even at a compatible pixel α will become very slim since

$$\text{Prob}(\Delta b_\alpha^n \neq 0) = \text{Prob}(\Delta b_\alpha^n = \pm 1) = \pm \text{E}[\Delta b_\alpha^n] = |\tau K^T [e^n]_\alpha| \leq \|e^n\|_\infty \ll 1,$$

following the proof of Theorem 4.1.

[3] Automatic exit for the algorithm could be based on, for example, the monitoring of the total squared errors $D(b^n \mid u) = \|e^n\|_2^2$. More quantitative discussion will be developed in the coming sections.

[4] In practice, the last line (iii-3) can be easily implemented via the simple random sampling (also see Shen [41]): take any random number f from the uniform distribution $U[0,1]$, and set b_α^{n+1} to 1 if $f \leq p_\alpha$, and to 0, otherwise.

[5] The major flipping step in (iii-3) is readily implementable in parallel machines.

It is also the ideal spot to explain why the above randomized algorithm has been entitled "Markovian." In fact, what the algorithm produces, is precisely a random walk through the (enormous) halftone image space $\{0,1\}^\Omega$:

$$b^0 \to b^1 \to \cdots \to b^n \to b^{n+1} \to \cdots .$$

It is indeed a Markov chain since

$$\text{Prob}\left(b^{n+1} \mid b^n, b^{n-1}, \cdots, b^0\right) = \text{Prob}\left(b^{n+1} \mid b^n\right). \tag{4.11}$$

This is plain to see from the above algorithm since the distribution of b^{n+1} has been completely determined by b^n (including the perceived error field $e^n = u - K[b^n]$).

In particular, unlike most conventional halftoning algorithms whose deterministic structures invariably lead to *fixed* final outputs, the LS-MGD algorithm seeks instead for optimal random fields. Pixelwise, the algorithm may produce *distinct* halftone images under different runs, which are however always similar, or *typical* in the stochastic or information-theoretic sense (see, e.g., Cover and Thomas [11]).

Next we investigate whether the LS-MGD algorithm indeed faithfully inherits the gradient descent feature from its deterministic counterparts (4.2) or (4.3). The two major theorems will crucially depend upon the *spatial smoothing* property of the HVS, which is first to be addressed.

5. Smoothing Degrees of the HVS: Entropy and Beyond. To expose most efficiently the very essence herein, assume for the moment that the HVS K is shift invariant and positive, with PSF k.

Earlier in Section 2, we have discussed that an HVS must be weakly lowpass:

$$K[1] \equiv 1, \qquad \text{or equivalently,} \qquad \sum_\alpha k_\alpha = 1. \tag{5.1}$$

This, however, should only be considered as a *minimum* requirement. For instance, if $K = Id$ is the identity operator, or equivalently, $k_\alpha = \delta_\alpha$ is the delta sequence, this minimum requirement is satisfied but obviously insufficient for modeling the HVS. The HVS *must* smoothen nearby information in order for a binary halftone image to be perceived as a continuum.

There are several alternatives for characterizing the degree of spatial smoothing. The classical approach turns to the spectral domain, and requires that the Fourier transform of $k = (k_\alpha)$ vanishes fast enough at high frequencies [33, 46]. A more direct characterization of spatial smoothing is inspired by the entropy measure in information theory and statistical mechanics [10, 11]. One can treat the positive PSF $k = (k_\alpha)$ as a probability distribution, and define its entropy by

$$S(k) = -\sum_\alpha k_\alpha \ln k_\alpha. \tag{5.2}$$

For instance, suppose k is supported on a local set I of $|I|$ pixels. Then

$$0 \leq S(k) \leq \ln |I|.$$

The lower bound is reached if and only if when k is supported on a single pixel (i.e., a delta measure), while the maximum upper bound is reached if and only if when k is *equally* distributed over all the pixels in I. Thus entropy provides a natural metric for measuring the degree of smoothing, which is precisely its innate physical meaning in statistical mechanics [10, 23].

Another simpler measure for characterizing spatial smoothing of a discrete probability sequence is the l^2 norm:

$$d = d(k) = \|k\|_2^2 = \sum_\alpha k_\alpha^2. \tag{5.3}$$

As for the entropy measure, assume that k is supported on a subset $I \subseteq \Omega$. Then one has

$$|I|^{-1} \leq d \leq 1,$$

since k is positive and satisfies the weak lowpass condition $\sum_\alpha k_\alpha = 1$. The upper bound is reached if and only if k is a delta-point source, while the lower bound is reached if and only k is equally distributed $k_\alpha \equiv |I|^{-1}$. Therefore,

$\frac{1}{d}$ *is proportional to the smoothing degree of the HVS.*

The inverse proportion between $S(k)$ and $d(k)$ is also partially manifest in the following inequality:

$$2S(k) = \sum_\alpha k_\alpha \ln k_\alpha^{-2} \geq \sum_\alpha k_\alpha \ln \|k\|_2^{-2} = \ln d(k)^{-1}.$$

Thus, a small d necessarily implies a large S, while both reflect good smoothing of the HVS. As an example, suppose that the support I of k is a 16 by 16 window, and k is equally distributed. Then

$$d = |I|^{-1} = (16 \times 16)^{-1} \leq 0.4\% \ll 1, \tag{5.4}$$

which is accordance with the ideal spatial smoothing.

Finally, for a general HSV K which is not necessarily shift invariant, we then define the following smoothing measure:

$$d = d(K) = \sup_{\alpha \in \Omega} \sum_{\beta \in \Omega} K_{\beta\alpha}^2 = \sup_{\alpha \in \Omega} \|K_{\cdot,\alpha}\|_2^2. \tag{5.5}$$

It is consistent with and thus generalizes the shift-invariant case. Moreover, we have the following.

PROPOSITION 5.1. *Consider K as a linear operator from $l^1(\Omega)$ to $l^2(\Omega)$. Then the above smoothing measure d is in fact a squared operator norm for the adjoint $K^T : l^2(\Omega) \to (l^1)^*(\Omega) = l^\infty(\Omega)$:*

$$d(K) = \|K^T\|_{l^2 \to l^\infty}^2.$$

Proof. By the definition of operator norm, one has

$$\|K^T\|_{l^2 \to l^\infty} = \sup_{v : \|v\|_2 \leq 1} \|K^T[v]\|_\infty.$$

For each pixel α,

$$|K^T[v]_\alpha\| = |\langle K_{\cdot,\alpha},\ v\rangle| \leq \|K_{\cdot,\alpha}\|_2 \cdot \|v\|_2 \leq \|K_{\cdot,\alpha}\|_2.$$

Therefore, for any v with $\|v\|_2 \leq 1$,

$$\|K^T[v]\|_\infty^2 = \sup_\alpha |K^T[v]_\alpha|^2 \leq \sup_\alpha \|K_{\cdot,\alpha}\|_2^2 = d(K),$$

implying that $\|K^T\|^2_{l^2\to l^\infty} \le d(K)$.

On the other hand, suppose at a pixel $\beta \in \Omega$:

$$\|K_{\cdot,\beta}\|_2^2 = \sup_\alpha \|K_{\cdot,\alpha}\|_2^2 = d(K).$$

Define $w = K_{\cdot,\beta}/\|K_{\cdot,\beta}\|_2$. Then,

$$|K^T[w]_\alpha| = |\langle K_{\cdot,\alpha},\ w\rangle| \le \|K_{\cdot,\alpha}\|_2\|w\|_2 = \|K_{\cdot,\alpha}\|_2 \le \sqrt{d(K)},$$

and both inequalities become equalities when $\alpha = \beta$. Thus $\|K^T[w]\|^2_\infty = d(K)$, implying that $d(K) \le \|K^T\|^2_{l^2\to l^\infty}$. (When $\#\Omega$ is infinite, the argument still holds by resorting to a maximizing sequence $\{\beta_n's\}$.) This completes the proof. □

As a result, a small $d(K)$ implies that $K^T : l^2 \to l^\infty$ tends to suppress isolated spikes in its output (since they alone can increase the l^∞-norm), and hence to encourage equal spreading or good spatial smoothing.

The following further characterizes $d(K)$ as the *perceptual sensitivity* to single-pixel flips.

PROPOSITION 5.2. *Given any halftone image b and any single pixel $\alpha \in \Omega$, define b^α to be the new halftone image which differs from b only at α:*

$$b^\alpha_\beta = \begin{cases} b_\beta, & \beta \ne \alpha, \\ 1 - b_\beta, & \beta = \alpha. \end{cases}$$

Then one has

$$d(K) = \sup_{\alpha\in\Omega, b\in\{0,1\}^\Omega} \|K[b^\alpha] - K[b]\|_2^2.$$

That is, $d(K)$ signifies the maximal (l^2) perceptual difference resulted from any localized single-pixel flip.

The proof is straightforward. Define $\delta^\alpha = (\delta^\alpha_\beta)_\beta$ to be the halftone image which is 1 only at $\beta = \alpha$, and always 0 at any other pixel $\beta \ne \alpha$. Then $K[b^\alpha] - K[b] = K[b^\alpha - b] = \pm K[\delta^\alpha] = \pm K_{\cdot,\alpha}$, and the proposition follows directly from the definition of $d(K)$.

This simple yet intriguing novel characterization reveals that a smaller $d(K)$ implies less perceptual sensitivity (by the HVS K) to single-pixel changes in any given halftone image b. In particular, a smaller $d(K)$ necessarily demands K to be less spatially localized or concentrated, which is consistent with the earlier picture that a smaller $d(K)$ signifies better spatial smoothing.

The important theoretical role of $d(K)$ in LS-MGD halftoning will be manifest in later analysis.

6. Gradient Descent Analysis for LS-MGD Halftoning. In this section, we show that the LS-MGD algorithm indeed decreases the total squared errors successively, which of course must be understood in the stochastic sense since the algorithm is essentially a Markov random walk.

We will establish two versions of essentially the same theorem. The first one is semi-heuristic but much easier to understand, while the second is mathematically rigorous but structurally more intricate. We shall always assume the validity of the compatibility condition at all pixels as discussed in Section 4, which substantially facilitates the analysis. Notice that rigorous analytical results as those developed herein have been very rare in the existing literature of halftoning.

6.1. Semi-Heuristic Version of Stochastic Gradient Descent. We first develop the semi-heuristic version of the gradient descent theorem, which is much easier to comprehend than the full and rigorous version in the next subsection, and yet efficiently reveals the very essence of the theorem.

At each time step n, we first define the conditional expectation operator E_n:

$$\mathrm{E}_n[X] = \mathrm{E}[X \mid b^n], \qquad \text{for any random variable } X. \tag{6.1}$$

Due to the stochastic nature of the LS-MGD algorithm, gradient descent is most naturally explored in the sense of such conditional expectations.

Suppose that the HSV is shift invariant with PSF k, and is sufficiently smoothing (as characterized by $d(k) \ll 1$ just discussed above). Then the *Law of Large Numbers* (LLN) leads to the following approximate identity [4]:

$$k * \Delta b^n_\alpha = \sum_\beta k_{\alpha-\beta} \Delta b^n_\beta \simeq \mathrm{E}_n[\Delta b^n_\alpha]. \tag{6.2}$$

The heuristic nature of this substitution lies in the assumptions that k is sufficiently smoothing, and that Δb^n_α's are independent and identically distributed (i.i.d.), in order to successfully invoke the LLN. (The next subsection shows more rigorously that the error of such substitution is controllable by the smoothing rate $d(k)$.) Consider the scenario given in (5.4) when the HSV k is equally distributed over a window I of 16 by 16. Then when Δb^n_α are i.i.d., one has,

$$k * \Delta b^n_\alpha = \frac{1}{|I|} \sum_{\beta \in \alpha - I} \Delta b^n_\beta = \frac{1}{256} \sum_{\beta \in \alpha - I} \Delta b^n_\beta,$$

whose mean is precisely $\mathrm{E}_n[\Delta b^n_\alpha]$, and variance is negligible since:

$$\mathrm{var}\,(k * \Delta b^n_\alpha) = \frac{1}{256} \mathrm{var}\,(\Delta b^n_\alpha) \leq \frac{1}{256} \times \frac{1}{4} = \frac{d(K)}{4} \leq 0.001.$$

Here $\mathrm{var}(\Delta b^n_\alpha) \leq 1/4$ since $\pm \Delta b^n_\alpha \in \{0, 1\}$ is Bernoulli (see Eqn. (4.6)). Thus in practice, such small variances make the heuristic substitution (6.2) sufficiently accurate. Generally, a small $d(K)$ or good smoothing would lend more justification to such a substitution. We are now ready to state the following heuristic but key observation.

Heuristic Observation on Gradient Descent. At each step n of the LS-MGD algorithm, define $e^n = u - k * b^n$ to be the *perceived* error field, and $D(b^n \mid u) = \|e^n\|_2^2$ the total squared errors, as in (3.1). Suppose the HSV k is positive and sufficiently smoothing so that the heuristic substitution (6.2) yields good approximation. Then the LS-MGD algorithm is gradient descent in the following stochastic sense:

$$\mathrm{E}_n[D(b^{n+1} \mid u)] - D(b^n \mid u) \simeq -\tau(2-\tau)\|k^T * e^n\|_2^2 \leq 0, \tag{6.3}$$

where $k^T_\alpha = k_{-\alpha}$ denotes the transpose, and $\tau \in (0, 1]$ is the marching size in the LS-MGD algorithm.

Proof. As before, we shall use the notation $\langle v,\ w \rangle = \sum_{\alpha \in \Omega} v_\alpha w_\alpha$ to denote the inner product in $l^2(\Omega)$. Since $b^{n+1} = b^n + \Delta b^n$, one has

$$\begin{aligned}
& D(b^{n+1} \mid u) - D(b^n \mid u) \\
&= \|u - k * b^{n+1}\|_2^2 - \|u - k * b^n\|_2^2 \\
&= \langle -k * b^{n+1} + k * b^n,\ 2u - k * (b^{n+1} + b^n) \rangle \\
&= -2\langle k * \Delta b^n,\ u - k * b^n - \frac{1}{2} k * \Delta b^n \rangle \\
&= -2\langle k * \Delta b^n,\ e^n \rangle + \langle k * \Delta b^n,\ k * \Delta b^n \rangle \\
&\simeq -2\langle \Delta b^n,\ k^T * e^n \rangle + \langle \mathrm{E}_n[\Delta b^n],\ \mathrm{E}_n[\Delta b^n] \rangle,
\end{aligned}$$

where only for the second part of the last equation the heuristic substitution (6.2) has bee invoked. Since e^n is deterministic with respect to the conditional expectation $\mathrm{E}_n[\cdot] = \mathrm{E}[\cdot \mid b^n]$, after the application of E_n, the last line continues:

$$\begin{aligned}
& -2\langle \mathrm{E}_n[\Delta b^n],\ k^T * e^n \rangle + \langle \mathrm{E}_n[\Delta b^n],\ \mathrm{E}_n[\Delta b^n] \rangle \\
&= -2\tau \langle k^T * e^n,\ k^T * e^n \rangle + \tau^2 \langle k^T * e^n,\ k^T * e^n \rangle \\
&= -\tau(2-\tau)\|k^T * e^n\|_2^2.
\end{aligned} \tag{6.4}$$

This completes the semi-heuristic "proof" of the observation. □

Before proceeding to the next subsection for the removal of the heuristics, we first explore the possible consequence of the established result in terms of convergence analysis.

If one removes the expectation operator E_n, as well as adopts the asymptotic limit (for the *Law of Large Numbers*), then the above heuristic observation *suggests* to investigate the decay behavior of a sequence of error fields $(e^n \mid n)$ that ideally satisfy:

$$\|e^{n+1}\|_2^2 - \|e^n\|_2^2 = -\tau(2-\tau)\|K^T[e^n]\|_2^2,$$

or more generally, in the inequality format,

$$\|e^{n+1}\|_2^2 - \|e^n\|_2^2 \le -\tau(2-\tau)\|K^T[e^n]\|_2^2, \qquad n = 0, 1, \cdots. \tag{6.5}$$

Suppose the lattice image domain Ω is finite. Then the linear HVS operator K is a finite matrix.

THEOREM 6.1 (Convergence under Ideal Iteration). *Suppose the HVS K is positive, weakly lowpass $K[1] \equiv 1$, and non-singular so that its minimum singular value $\sigma = \sigma_{\min}$ is positive. Define for $\tau \in (0, 1]$,*

$$\rho = \tau(2-\tau)\sigma^2 > 0.$$

Then the ideal *iteration formula* (6.5) *implies a linear convergence rate for the error fields:*

$$\|e^n\|_2^2 \le (1-\rho)^n \|e^0\|_2^2, \quad n = 0, 1, \cdots.$$

Proof. By the definition of singular values, for any $w \in l^2(\Omega)$, one has,

$$\sigma_{\min}^2 \|w\|_2^2 \le \|K^T[w]\|_2^2 = \langle KK^T[w],\ w\rangle \le \sigma_{\max}^2 \|w\|_2^2.$$

(Notice that K and K^T share the same set of singular values.) Then the ideal iteration formula (6.5) implies that

$$\|e^{n+1}\|_2^2 \le \|e^n\|_2^2 - \tau(2-\tau)\sigma^2\|e^n\|_2^2 = (1-\rho)\|e^n\|_2^2.$$

Iterating this formula leads to $\|e^n\|_2^2 \le (1-\rho)^n\|e^0\|_2^2$.

Finally we point out that ρ indeed falls into the range of $(0, 1]$ for any marching step $\tau \in (0, 1]$ so that $1 - \rho \in [0, 1)$. Let $\|K\|_F$ denote the Fröbenius norm $\sqrt{\mathrm{trace}(KK^T)}$, and $N = |\Omega|$. Since K is positive and satisfies the weakly lowpass condition $K[1] \equiv 1$, one has

$$\begin{aligned} N\sigma^2 &\le \sigma_1^2 + \cdots + \sigma_N^2 = \|K\|_F^2 \\ &= \sum_{\alpha\in\Omega}\sum_{\beta\in\Omega} K_{\alpha\beta}^2 \le \sum_{\alpha\in\Omega}\left(\sum_{\beta\in\Omega} K_{\alpha\beta}\right)^2 \\ &= \sum_{\alpha\in\Omega} 1 = N. \end{aligned}$$

Thus one must have $\sigma^2 \le 1$ and $\rho = \tau(2-\tau)\sigma^2 \in (0, 1]$ when $\tau \in (0, 1]$, since $\tau(2-\tau)$ reaches its peak value 1 at $\tau = 1$. □

As an example, consider a 1-dimensional image array with only 4 pixels. Suppose the human vision system is specified by the following 4 by 4 matrix:

$$K = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}.$$

Then the smallest singular value $\sigma = \sigma_{\min} = 0.2185...$ is indeed positive.

Notice that Theorem 6.1 would imply the absolute convergence $e^n \to 0$ as $n \to \infty$, which of course could only be true under the *idealistic* control formula (6.5). In reality, the stochastic nature of the algorithm can never yield such simple and clean error control, and the expectation operator must be included as in the preceding Heuristic Observation, and the error due to the heuristic LLN substitution also must be investigated. Through the analysis of such idealistic error control, however, the above theorem does effectively provide a simple mechanism for explaining the robust iterative behavior of the LS-MGD algorithm, as observed in all the computational examples later on (esp. Figure 7.3).

6.2. Mathematically Rigorous Version of Stochastic Gradient Descent. In this subsection, we will treat the above heuristic LLN substitution with mathematical rigor, which however leads to a much more technical theorem. To allow finite lattice domains, we shall assume that the HSV K is linear, but not necessarily shift invariant.

Following the LS-MGD algorithm, at each step n, define the *mean flip field* F by (with the superscript n omitted):

$$F = \mathrm{E}_n[\Delta b^n] = \tau\, K^T[e^n], \quad \text{or} \quad F_\alpha = \mathrm{E}_n[\Delta b^n_\alpha] = \tau\, K^T[e^n]_\alpha, \quad \alpha \in \Omega.$$

THEOREM 6.2 (Stochastic Gradient Descent). *Let $N = |\Omega|$ be the number of pixels of the image lattice Ω, and $d = d(K)$ be defined in (5.5) for measuring the smoothing degree of the HVS. Assume that the HVS K is positive (i.e., $K_{\alpha\beta} \geq 0$ for any $\alpha, \beta \in \Omega$) and symmetric: $K^T = K$, and satisfies the weak lowpass condition $K[1] \equiv 1$. Then at each step n of the LS-MGD marching, one has*

$$\mathrm{E}_n[D(b^{n+1} \mid u)] - D(b^n \mid u) \leq -(2\tau^{-1} - 1 + d)\|F\|_2^2 + d\sqrt{N}\, \|F\|_2.$$

In particular, stochastic gradient descent $\mathrm{E}_n[D(b^{n+1} \mid u)] < D(b^n \mid u)$ is achieved for any $\varepsilon > 0$ and any marching step size τ with

$$\tau \leq \tau_\varepsilon = \min\left(\frac{2\varepsilon}{d\sqrt{N} + (1-d)\varepsilon},\ 1\right), \tag{6.6}$$

as long as the mean flip field at step n still remains large with respect to ε: $\|F\|_2 > \varepsilon$.

Proof. As in the first half of the proof for the Heuristic Observation in the preceding section, one has

$$D(b^{n+1} \mid u) - D(b^n \mid u) = -2\langle K[\Delta b^n],\ e^n\rangle + \langle K[\Delta b^n],\ K[\Delta b^n]\rangle. \tag{6.7}$$

The second term on the right can be rewritten as:

$$\begin{aligned}\langle K[\Delta b^n],\ K[\Delta b^n]\rangle &= \langle K^T K[\Delta b^n],\ \Delta b^n\rangle \\ &= \sum_{\alpha,\beta} [K^T K]_{\alpha\beta} \Delta b^n_\alpha \Delta b^n_\beta \\ &= \sum_{\alpha\neq\beta} [K^T K]_{\alpha\beta} \Delta b^n_\alpha \Delta b^n_\beta + \sum_\alpha [K^T K]_{\alpha\alpha} \left(\Delta b^n_\alpha\right)^2.\end{aligned} \tag{6.8}$$

Since the LS-MGD algorithm is carried out *pixelwise* via:

$$\pm\Delta b^n_\alpha \sim \mathrm{Bernoulli}(F_\alpha), \qquad \alpha \in \Omega,$$

Δb^n_α and Δb^n_β must be *independent* for any two distinct pixels α and β, and as a result,

$$\mathrm{E}_n[\Delta b^n_\alpha \Delta b^n_\beta] = \mathrm{E}_n[\Delta b^n_\alpha] \times \mathrm{E}_n[\Delta b^n_\beta], \qquad \alpha \neq \beta.$$

Thus after the application of E_n, the last equation in (6.8) continues as follows:

$$\begin{aligned}&\sum_{\alpha\neq\beta} [K^T K]_{\alpha\beta} \mathrm{E}_n[\Delta b^n_\alpha]\ \mathrm{E}_n[\Delta b^n_\beta] + \sum_\alpha [K^T K]_{\alpha\alpha} \mathrm{E}_n\left[\left(\Delta b^n_\alpha\right)^2\right] \\ &= \sum_{\alpha,\beta} [K^T K]_{\alpha\beta} \mathrm{E}_n[\Delta b^n_\alpha]\ \mathrm{E}_n[\Delta b^n_\beta] + \sum_\alpha [K^T K]_{\alpha\alpha} \left(\mathrm{E}_n\left[\left(\Delta b^n_\alpha\right)^2\right] - \left(\mathrm{E}_n[\Delta b^n_\alpha]\right)^2\right) \\ &= \sum_{\alpha,\beta} [K^T K]_{\alpha\beta} F_\alpha F_\beta + \sum_\alpha [K^T K]_{\alpha\alpha} V_\alpha = \|K[F]\|_2^2 + \sum_\alpha [K^T K]_{\alpha\alpha} V_\alpha,\end{aligned} \tag{6.9}$$

where $F = \mathrm{E}_n[\Delta b^n] = \tau K^T[e^n]$ is the mean flip field defined earlier, and

$$V = (V_\alpha \mid \alpha \in \Omega): \quad V_\alpha = \mathrm{var}\left(\Delta b^n_\alpha\right) = \mathrm{E}_n\left[\left(\Delta b^n_\alpha\right)^2\right] - \left(\mathrm{E}_n[\Delta b^n_\alpha]\right)^2,$$

is the variance scalar field for the flip random field Δb^n. By the definition of the smoothing measure d (5.5), one has

$$[K^T K]_{\alpha\alpha} = \sum_\beta K^T_{\alpha\beta} K_{\beta\alpha} = \sum_\beta K^2_{\beta\alpha} \leq d(K).$$

On the other hand, by the discrete Young's inequality [28], one has for any $w \in l^2(\Omega)$,

$$\|K[w]\|_2 \leq \left(\sup_\alpha \sum_\beta |K_{\beta\alpha}|\right)^{1/2} \left(\sup_\beta \sum_\alpha |K_{\beta\alpha}|\right)^{1/2} \|w\|_2.$$

Since K has been assumed to be positive, symmetric, and satisfies $K[1] = 1$, the two square root factors on the right hand side are both 1. Then $\|K[w]\|_2 \leq \|w\|_2$. Thus the last equation of (6.9) leads to:

$$\mathrm{E}_n\left[\langle K[\Delta b^n],\ K[\Delta b^n]\rangle\right] \leq \|F\|_2^2 + d(K)\sum_\alpha V_\alpha = \|F\|_2^2 + d(K)\|V\|_1. \tag{6.10}$$

Therefore, combined with (6.7), this leads to

$$\begin{aligned} \mathrm{E}_n\left[D(b^{n+1} \mid u)\right] - D(b^n \mid u) &\leq -2\langle K\left[\mathrm{E}_n[\Delta b^n]\right],\ e^n\rangle + \|F\|_2^2 + d\|V\|_1 \\ &= -2\langle F,\ K^T[e^n]\rangle + \|F\|_2^2 + d\|V\|_1 \\ &= -(2\tau^{-1} - 1)\|F\|_2^2 + d\|V\|_1. \end{aligned} \tag{6.11}$$

We now work on the variance scalar field V. Generally suppose $X \sim \mathrm{Bernoulli}(p)$, then

$$p = \mathrm{E}[X], \qquad \mathrm{var}(X) = p(1-p) = \mathrm{E}[X] - \mathrm{E}[X]^2.$$

As a result, more generally if $\pm X \sim \mathrm{Bernoulli}(p)$, one must have

$$\mathrm{var}(X) = |\mathrm{E}[X]| - \mathrm{E}[X]^2.$$

Applying this to $X = \Delta b^n_\alpha$ pixelwise, one has

$$V_\alpha = |F_\alpha| - F_\alpha^2, \qquad \|V\|_1 = \|F\|_1 - \|F\|_2^2.$$

Then following (6.11), one has

$$\mathrm{E}_n\left[D(b^{n+1} \mid u)\right] - D(b^n \mid u) \leq -(2\tau^{-1} - 1 + d)\|F\|_2^2 + d\|F\|_1.$$

By Schwarz's inequality, one has,

$$\|F\|_1 \leq \|1\|_2\ \|F\|_2 = \sqrt{|\Omega|}\ \|F\|_2 = \sqrt{N}\ \|F\|_2.$$

As a result,

$$\begin{aligned} &\mathrm{E}_n\left[D(b^{n+1} \mid u)\right] - D(b^n \mid u) \\ &\leq -(2\tau^{-1} - 1 + d)\|F\|_2^2 + d\sqrt{N}\ \|F\|_2 \\ &\leq -\left((2\tau^{-1} - 1 + d)\|F\|_2 - d\sqrt{N}\right) \times \|F\|_2. \end{aligned} \tag{6.12}$$

The expression for the critical τ_ε in (6.6) is then obtained by setting

$$(2\tau_\varepsilon^{-1} - 1 + d)\varepsilon - d\sqrt{N} = 0,$$

and by noticing that $\tau \leq 1$ as established in Theorem 4.1. This completes the proof. □

Notice that the much simpler semi-heuristic theorem in the preceding subsection corresponds precisely to the asymptotic limit of the current theorem when the smoothing measure $d \to 0$ and the upper bound $\tau_\varepsilon \to 1$, which is independent of ε or $\|F\|_2$.

7. Computational Examples and Performance Monitoring. In this section, through typical test images, we discuss the computational performance and convergence behavior of the proposed LS-MGD halftoning algorithm. In all the examples, the marching step size τ is either $\tau = 1$ or $\tau = 1/2$ (see Theorem 4.1), and the HVS K is modelled by the canonical isotropic Gaussian (see Eqn. (2.2)).

First displayed in Figure 7.1 are the halftone outputs from three different methods for a stair-ramp test image $u(x, y)$ given by:

$$u(x,y) = \begin{cases} 1 - \frac{2}{3}x, & \text{for} \quad x \in [0, 1/2], \\ \frac{2}{3}(1-x), & \text{for} \quad x \in (1/2, 1]. \end{cases}$$

The three halftone images are produced separately by Floyd and Steinberg's classic error diffusion, Perona-Malik's error diffusion and stochastic flipping (PM-SF) [41, 39], and the LS-MGD algorithm proposed in the present work. Notice the most striking difference between the first method and the other two: the first is a deterministic process, while the other two are both stochastic and output random fields eventually in stochastic equilibrium. It is well known that Floyd and Steinberg's algorithm may produce visible deterministic artifacts (e.g., "worms" or "tearing") for certain class of images, while the latter two have both avoided such deficiency due to the stochastic nature.

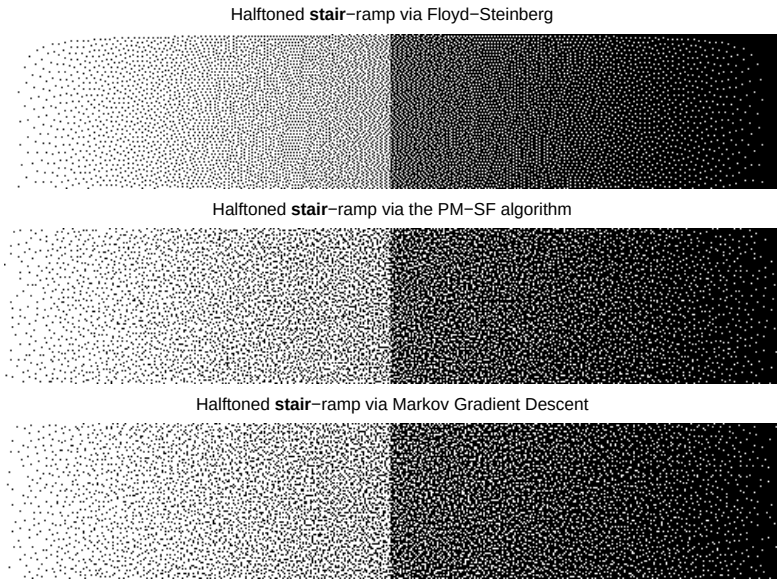

FIG. 7.1. *Comparing three different approaches to halftoning a stair-ramp test image: Floyd-Steinberg's classic error diffusion [18] (top), Perona-Malik error diffusion and stochastic flipping (PS-SF) [41] (middle), and the LS-MGD method proposed in the current work (bottom). One can observe the clear difference between the deterministic nature of Floyd-Steinberg's algorithm and the stochastic (or random-field) nature of the latter two.*

Figure 7.2 shows the LS-MGD output of the Barbara test image. It has been displayed at about 100 dots per inch (dpi). (Most commercial printers have about $O(10^3)$ dpi's.) Notice that the line textures on the scarf and the table cloth have been resolved sufficiently well even at such a low dpi rate.

To further objectively monitor the random walk of LS-MGD halftoning, we introduce two key quantities $\text{frpp}(n)$ and $\text{psepp}(n)$ at each step n.

The *flip rate per pixel* frpp was first introduced in the earlier work [41] to quantify the flipping activity at each step, and is defined by

$$\text{frpp}(n) = \frac{1}{|\Omega|} \sum_{\alpha \in \Omega} |b_\alpha^{n+1} - b_\alpha^n| = \frac{1}{|\Omega|} \|\Delta b^n\|_1. \tag{7.1}$$

That is, $|\Omega|\text{frpp}(n)$ denotes the total number of pixels at which the halftones are flipped to their opposites at step n. Thus statistically, $\text{frpp}(n)$ could be understood as the likelihood at step n that any single pixel will

Halftone Barbara Image via MGD

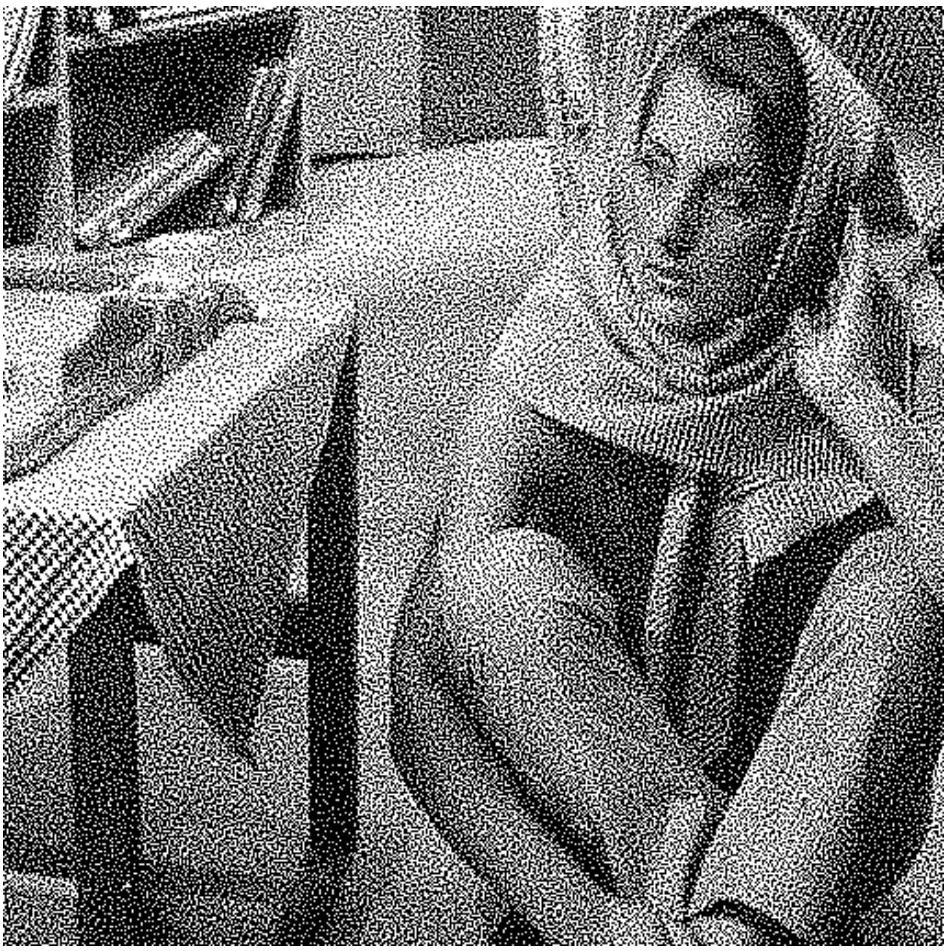

FIG. 7.2. *Halftoned Barbara based on the LS-MGD model and algorithm (at about 100 dpi).*

be flipped to its opposite. Since the algorithm is in essence a Markov random walk, absolute convergence with $\mathrm{frpp}(n) = 0$ on a large domain is very unlikely, and statistical convergence or equilibrium must be understood in the sense of a small and stable $\mathrm{frpp}(n)$. Plotted on the middle panel of Figure 7.3 is the frpp curve for LS-MGD halftoning of the pepper image on the left.

The second quantity $\mathrm{psepp}(n)$, the *perceived square error per pixel* at each step n, has been motivated by Theorem 6.2, and is defined by

$$\mathrm{psepp}(n) = \frac{1}{|\Omega|}\sum_{\alpha\in\Omega}(e^n_\alpha)^2 = \frac{1}{|\Omega|}\|e^n\|_2^2 = \frac{1}{|\Omega|}D(b^n \mid u). \tag{7.2}$$

Since $(b^n \mid n = 0, 1, \cdots)$ is a random walk, $\mathrm{psepp}(n)$ strictly speaking must be a random sequence as well. As theoretically predicted by both the Heuristic Observation and 6.2, one can expect that $\mathrm{psepp}(n)$ should decrease *on average*, at least in terms of conditional expectations. The most striking observation, as shown on the lower left panel of Figure 7.3, is that the sequence $\mathrm{psepp}(n)$ *actually* monotonically decreases without any stochastic fluctuation! To view the numerical details of convergence, we have also shown the semilog plot of the normalized $\mathrm{psepp}(n)$ (on the lower right panel in Figure 7.3):

$$\delta(n) = \log_{10}(\mathrm{psepp}(n) - p_*), \qquad n = 1, 2, \cdots,$$

where $p_* = \lim_n \mathrm{psepp}(n)$. Since for a fixed HVS K and a generic contone image u, the equation $K[b] = u$

typically has no halftone solution b, one concludes that generically p_* must be nonzero (see Figure 7.3). In practice, p_* can be approximated by the last psepp(N) for an almost converged iteration scheme $n = 1 : N$. The almost linear behavior on the lower right panel suggests a linear convergence rate in the form of:

$$\text{psepp}(n) \approx p_* + a\rho^n, \qquad \text{for some } a > 0 \text{ and } \rho \in (0, 1).$$

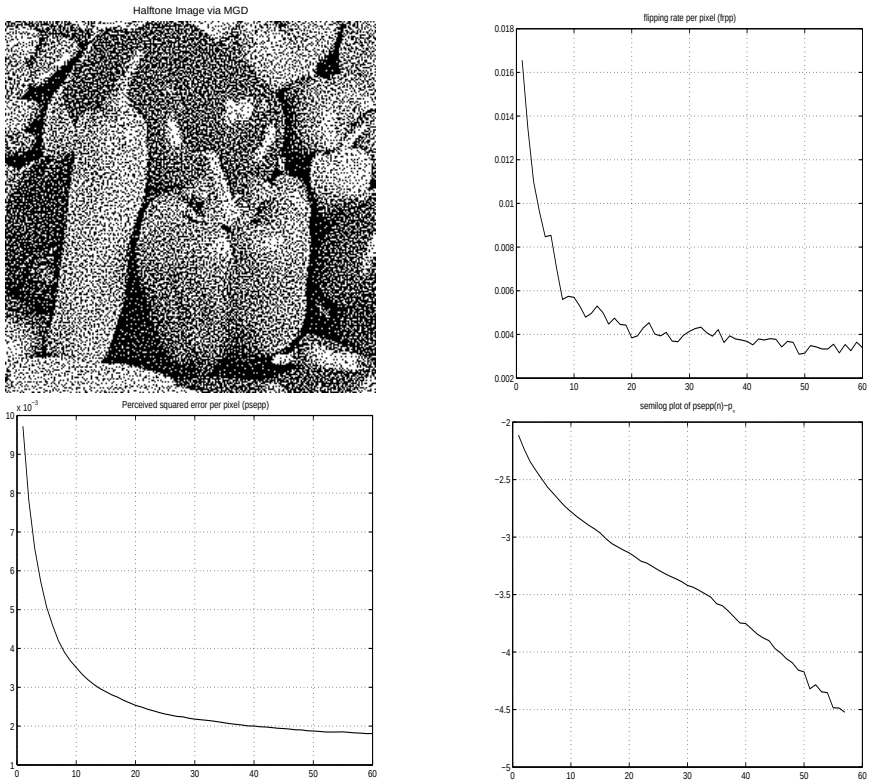

FIG. 7.3. *Monitoring the LS-MGD algorithm: Halftoned peppers (upper left), the curve of* flipping rate per pixel *frpp(n) at each step n defined in Eqn.* (7.1) *(upper right), and the curve of* perceived square error per pixel *psepp(n) at each step n defined in Eqn.* (7.2) *(lower left), and its semilog plot* $\log_{10}(\text{psepp}(n) - p_*)$ *(lower right) with* p_* *denoting the limiting value. An* frpp(n) *of* 0.004*, for example, means that* on average b^{n+1} *differs from* b^n *only on 4 pixels within any* 32×32 *window. Due to the stochastic nature of the LS-MGD algorithm, absolute convergence with a zero flipping rate should not be expected. The last plot hints an approximately linear convergence rate.*

Following the suggestion of one of our referees, we herein also compare *quantitatively* the performance of the LS-MGD method with that of the Floyd-Steinberg algorithm. Given a human vision system model K, we monitor the final square error $D(b \mid u)$ for both methods, or equivalently, the psepp $D(b \mid u)/|\Omega|$.

LS-MGD specifically targets at minimizing $D(b \mid u)$, but only in an approximate and stochastic sense in order to reduce the enormous burden of searching through the *entire* halftone space of astronomical dimension $2^{|\Omega|}$. The built-in stochastic/thermal noise thus keeps LS-MGD away from reaching the absolute minimum of $D(b \mid u)$. The Floyd-Steinberg algorithm, on the other hand, adopts a fixed routine that incorporates *neither* the human vision system K *nor* the sampling rate of the given image lattice. It is a purely lattice-based approach, and proceeds in exactly the same manner for two contone images sampled at different spatial rates or a same image perceived in different distances/scales. As a result it has no built-in scale adaptivity.

In light of all these qualitative observations, we can understand the numerics gathered in Table 7.1. The test image is the Pepper image in 512×512, and the HVS system K is modelled by an isotropic 2-D Gaussian with variance σ^2. For convenience, we shall simply call σ the *scale* of the HVS. From a given reference numerical scale σ_0, we have compared the two models in three different scales: $\sigma = \sigma_0$, $\sqrt{2}\sigma_0$, and $\sqrt{3}\sigma_0$.

Table 7.1
Comparing the psepp's of LS-MGD and FS under different HVS scales.

HVS Scale	$\sqrt{3}\sigma_0$	$\sqrt{2}\sigma_0$	σ_0
LS-MGD psepp	0.0030	0.0027	0.0028
FS psepp	0.0036	0.0030	0.0023
FS Percentage Higher	20%	11%	-18%

The Floyd-Steinberg algorithm is strictly based on a *local* balancing scheme (via error diffusion into direct neighbors in the image lattice), and thus performs well for an HVS K with a smaller scale $\sigma = \sigma_0$ even though it has no knowledge of the HVS itself. The perceived square error per pixel psepp is about -18% below the LS-MGD method. As the scale σ increases, however, the Floyd-Steinberg algorithm starts to suffer from its ignorance of the HVS and its lack of scale-adaptivity. The psepp starts to exceed that of the LS-MGD method by 11% when $\sigma = \sqrt{2}\sigma_0$, and by 20% when $\sigma = \sqrt{3}\sigma_0$.

Next, to better visualize the random walk during LS-MGD halftoning, in Figure 7.4, we have displayed the intermediate halftone images b^n for $n = 0, 6, 12$, and 18. The differences among $n = 0, 6$, and 12 can be easily perceived. The perceptual difference between $n = 12$ and $n = 18$ is much less apparent, but can be readily discerned from the monitoring curves like those in the preceding figure.

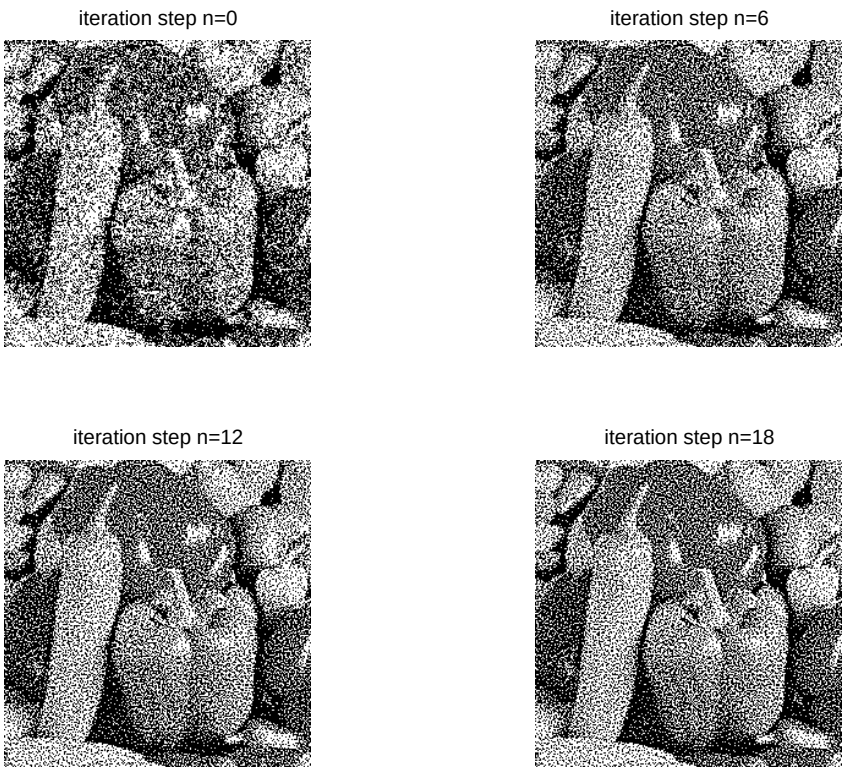

Fig. 7.4. *Halftoned peppers via LS-MGD halftoning at four different stages of the Markov random walk:* $n = 0, 6, 12,$ *and* 18.

Finally, recall in signal processing [33] that a *white* noise is a random signal or image v whose power spectral density [33] $P_{vv}(\omega) = \sigma^2$ is constant, or equally spreads over *all* frequencies $\omega \in [-\pi, \pi]$ or $[-\pi, \pi]^2$. A *blue* noise is, on the other hand, a random signal or image v whose spectral powers $P_{vv}(\omega)$ are concentrated on the blue band or high frequencies.

As discovered in the classic work by Ulichney [48, 49], a perceptually satisfactory halftone image must demonstrate the character of a blue-noise field. In Figure 7.5, we have plotted the *normalized* power spectral densities $P_{vv}(\Omega)$ for the halftone images b corresponding to a constant image $u \equiv 0.35$, obtained separately

by Floyd and Steinberg's error diffusion (left panel) and by the LS-MGD proposed in the present work (right panel). Here the normalization shifts the mean to zero: $v = b - 0.35$, so that the sharp δ-spike at the zero-frequency center has been tamed.

As manifested in Figure 7.5, compared with the power spectra of Floyd-Steinberg halftoning, those of the LS-MGD are more *uniformly* distributed in the blue (or high-frequency) band, and the distribution is almost rotationally invariant. Also notice that the power spectra from Floyd-Steinberg halftoning carry visible line structures and hence introduce undesirable biases favoring special directions.

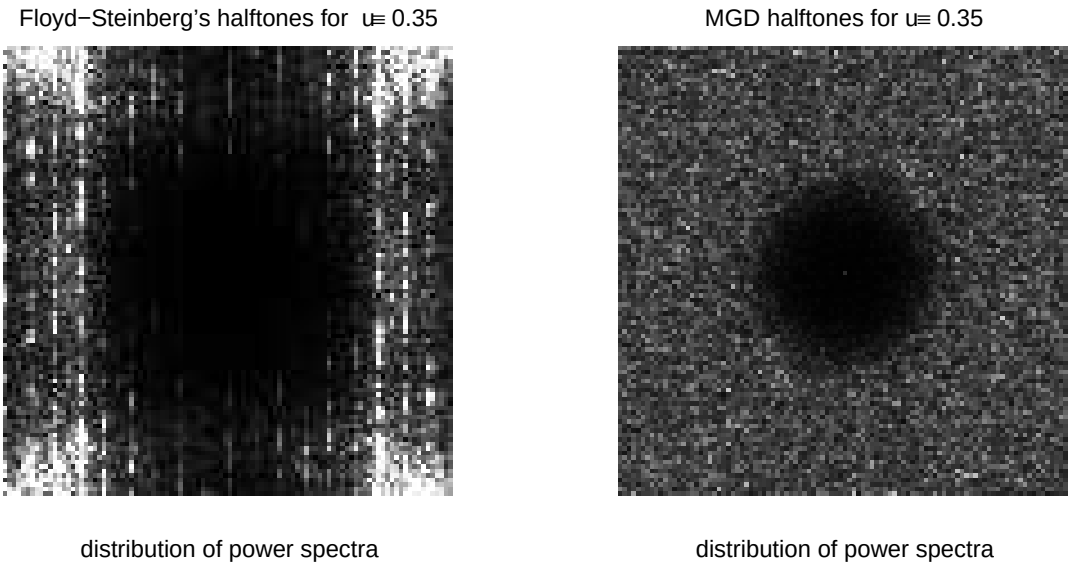

FIG. 7.5. *The blue-noise character of the LS-MGD halftones of a constant image $u \equiv 0.35$: compared with the power spectra of Floyd-Steinberg's halftoning (left panel), the spectral powers of the LS-MGD halftoning (right panel) are more uniformly distributed in the blue (or high-frequency) band, and the distribution appears more isotropic or rotationally invariant. (Displayed on each panel is the power spectral density $P_{vv}(\omega_1, \omega_2)$ for the mean-shifted tone field $v = b - 0.35$, which has zero mean.)*

Acknowledgments. The author wishes to thank Professors Ingrid Daubechies, Ron DeVore, Sinan Güntürk, Eitan Tadmor, and John Benedetto for their generous support and suggestions on the current project. The author is also enormously indebted to the teaching and encouragement from two industrial leaders on halftoning: Dr. Chai Wah Wu at IBM and Dr. Gabriel Marcu at the Apple Computer, Inc. Without the professional insights or the generous help from the referees, the revision effort could not have possibly resulted in a much polished manuscript. Finally, the author would like to dedicate this work to his dear friends who have been generously and unconditionally supporting him throughout the years: Professors Gil Strang, Tony Chan, Stan Osher, Andrea Bertozzi, Luminita Vese, Steve Smale, Haomin Zhou, Yang Wang, Ding-Xuan Zhou, Raymond Chan, Sung-Ha Kang, Yoon-Mo Jung, Jin-Keun Seo, and Kevin Vixie.

REFERENCES

[1] Z. Baharav and D. Shaked. Watermarking of dither halftoned images. *Proc. SPIE*, pages 307–313, 1999.
[2] F. A. Baqai and J. P. Allebach. Halftoning via direct binary search using analytical and stochastic printer models. *IEEE Trans. Image Processing*, 12(1):1–15, 2003.
[3] B. E. Bayer. An optimum method for two level rendition of continuous-tone pictures. *Proc. IEEE Int'l Conf. Communication*, 26:11–15, 1973.
[4] P. Billingsley. *Probability and Measure.* Wiley-Interscience Publisher, 1995.
[5] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.
[6] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.

[7] T. F. Chan and J. Shen. *Image Processing and Analysis: variational, PDE, wavelet, and stochastic methods.* SIAM Publisher, Philadelphia, 2005.
[8] T. F. Chan and J. Shen. Variational image inpainting. *Comm. Pure Applied Math.*, 58:579–619, 2005.
[9] T. F. Chan and J. Shen. Theory and computation of variational image deblurring. In *IMS (Inst. Math. Sci., Singapore) Lecture Notes on Mathematics and Computation in Imaging Science and Information Processing.* World Scientific Publishing Co., 2007.
[10] D. Chandler. *Introduction to Modern Statistical Mechanics.* Oxford University Press, New York and Oxford, 1987.
[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., New York, 1991.
[12] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. AMS*, 39(1):1–49, 2001.
[13] Z. Cvetcovic, I. Daubechies, and B. Logan. Single-bit oversampled A/D conversion with exponential accuracy in the bit-rate. *IEEE Trans. Inf. Theory,* submitted, 2005.
[14] I. Daubechies and R. DeVore. Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order. *Annals of Math.*, 158(2):679–710, 2003.
[15] I. Daubechies, R. DeVore, C. S. Güntürk, and V.A. Vaishampayan. Beta expansions: A new approach to digitally correct A/D conversion. *Proc. IEEE Int'l Symp. Circ. Sys. (ISCAS'2002), Scottsdale, Arizona*, pages 26–29, 2002.
[16] R. Eschbach. Error diffusion algorithm with homogeneous response in highlight and shadow areas. *J. Electronic Imaging*, 6:348–356, 1997.
[17] Z. Fan and R. Eschbach. Limit cycle behavior of error diffusion. *Proc. IEEE Int. Conf. Image Processing*, 2:1041–1045, 1994.
[18] R. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. *Proc. Soc. Info. Disp.*, 17(2):75–77, 1976.
[19] B. Frieden. Aberration reduction by multiple relays of an incoherent image. *J. Opt. Soc. Am. (A)*, 21:1834–1840, 2004.
[20] M. S. Fu and O. C. Au. Data hiding watermarking for halftone images. *IEEE Trans. Image Processing*, 11(4):477–484, 2002.
[21] R. Geist, R. Reynolds, and D. Suggs. A Markovian framework for digital halftoning. *ACM Trans. Graphics*, 12(2):136–159, 1993.
[22] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.
[23] W. Gibbs. *Elementary Principles of Statistical Mechanics.* Yale University Press, 1902.
[24] C. S. Güntürk. *Harmonic Analysis of Two Problems in Signal Quantization and Compression.* PhD thesis, Princeton University, 2000.
[25] C. S. Güntürk and N. T. Thao. Refined error analysis in second-order Sigma-Delta modulation with constant inputs. *IEEE Trans. Inform. Theory*, 50(5):839–860, 2004.
[26] D. E. Knuth. Digital halftones by dot diffusion. *ACM Trans. Graph.*, 6:245–273, 1987.
[27] P. Li and J. P. Allebach. Tone-dependent error diffusion. *IEEE Trans. Image Processing*, 13(2):201–215, 2004.
[28] E. H. Lieb and M. Loss. *Analysis.* Amer. Math. Soc., second edition, 2001.
[29] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.
[30] M. Mese and P. P. Vaidyanathan. Optimized halftoning using dot diffusion and methods for inverse halftoning. *IEEE Trans. Image Processing*, 9(4):691–709, 2000.
[31] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.
[32] D. L. Neuhoff, T. N. Pappas, and N. Seshadri. One-dimensional least-squares model-based halftoning. *Proc. ICASSP-92, San Francisco, CA*, pages 189–192, 1992.
[33] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice Hall Inc., New Jersey, 1989.
[34] D. Ozdemir and L. Akarun. Fuzzy algorithms for combined quantization and dithering. *IEEE Trans. Image Processing*, 10(6):923–931, 2001.
[35] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
[36] J. Shen. Inpainting and the fundamental problem of image processing. *SIAM News*, 36, 2003.
[37] J. Shen. On the foundations of vision modeling I. Weber's law and Weberized TV (total variation) restoration. *Physica D: Nonlinear Phenomena*, 175:241–251, 2003.
[38] J. Shen. Bayesian video dejittering by BV image model. *SIAM J. Appl. Math.*, 64(5):1691–1708, 2004.
[39] J. Shen. Halftoning via Perona-Malik diffusion and stochastic flipping (invited letter). *SPIE's Electronic Imaging Newsletter*, 16(2):6, 2005.
[40] J. Shen. Piecewise $H^{-1} + H^0 + H^1$ images and the Mumford-Shah-Sobolev model for segmented image decomposition. *Appl. Math. Res. Exp.*, 4:143–167, 2005.
[41] J. Shen. Progressive halftoning by Perona-Malik error diffusion and stochastic flipping. *Proc. SPIE*, 6064A(3), 2005.
[42] J. Shen and Y.-M. Jung. Weberized Mumford-Shah model with Bose-Einstein photon noise. *Appl. Math. Optim.*, 53(3):331–358, 2006.
[43] S. Smale and D.-X. Zhou. Shannon sampling and function reconstruction from point values. *Bull. Amer. Math. Soc.*, 41:279–305, 2004.
[44] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, MA, 1993.
[45] G. Strang. *Introduction to Linear Algebra.* Wellesley-Cambridge Press, 3rd edition, 1998.
[46] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.
[47] J. Sullivan, R. Miller, and G. Pios. Image halftoning using a visual model in error diffusion. *J. Opt. Soc. Amer. A*, 10(8):1714–1724, 1993.
[48] R. Ulichney. *Digital Halftoning.* MIT Press, Cambridge, Massachusetts, 1987.

[49] R. Ulichney. Dithering with blue noise. *Proc. IEEE*, 76:56–79, 1988.
[50] A. Vongkunghae, J. Yi, and R. B. Wells. A printer model using signal processing techniques. *IEEE Trans. Image Processing*, 12(7):776–783, 2003.
[51] I. H. Witten and R. M. Neal. Using Peano curves for bilevel display of continuous-tone images. *IEEE Comput. Graph. Applicat.*, 2:47–52, 1982.
[52] P. W. Wong. Adaptive error diffusion and its application in multiresolution rendering. *IEEE Trans. Image Processing*, 5:1184–1196, 1996.
[53] P. W. Wong and J. P. Allebach. Optimum error diffusion kernel design. *Proc. SPIE*, 3018:236–242, 1997.