Nonnegative Definite Quadratic Penalty Design for Penalized-Likelihood Reconstruction

J. Webster Stayman and Jeffrey A. Fessler
EECS Department, University of Michigan

Abstract—**Likelihood-based estimators with conventional regularization methods generally produces images with nonuniform and anisotropic spatial resolution properties. Previous work on penalty design for penalized-likelihood estimators has led to statistical reconstruction methods that yield approximately uniform "average" resolution. However some asymmetries in the local point-spread functions persist. Such anisotropies result in the elongation of otherwise symmetric features like circular lesions. All previously published penalty functions have used nonnegative values for the weighting coefficients between neighboring voxels. Such nonnegativity provides a sufficient (but not necessary) condition to ensure that the penalty function is convex, which in turn ensures that the objective function has a unique maximizer. This paper describes a novel method for penalty design that allows a subset of the weighting coefficients to take negative values, while still ensuring convexity of the penalty function. We demonstrate that penalties designed under these more flexible constraints yield local point-spread functions that are more isotropic than the previous penalty design methods for 2D PET image reconstruction.**

I. Motivation

Statistical reconstruction methods are often preferred over analytic methods because they generally yield lower noise images, and allow for the incorporation of a wide range of realistic system effects (*i.e.*: nonuniform attenuation, nonuniform sampling, etc.). However, statistical methods rarely yield images with uniform resolution properties.[1] For example, it is well known in systems that possess a shift-variant geometric response (as in SPECT where detectors have a depth-dependent response, or in PET where nonuniform sampling or depth-of-interaction effects exist), penalized-likelihood estimators yield images with nonuniform resolution properties. Furthermore nonuniformities will exist even in shift-invariant systems due to the implicit data weighting by the estimator. [1] Such nonuniformities can lead to shape distortion when resolution properties are anisotropic, and can complicate both qualitative and quantitative comparisons between images or image regions.

In contrast, analytic methods like filtered backprojection (FBP) yield uniform resolution properties for systems with a uniform geometric response. Or, for nonuniform systems like SPECT, analytic methods have been developed that compensate for effects like the depth-dependent response [2, 3]. However, such methods generally ignore the noise model, or rely on approximations to the actual system model so that nonuniformities cannot be fully compensated. This is our motivation to modify a statistical approach to provide both uniform resolution properties and good noise performance. It has been our intention to use a shift-variant regularization in a penalized-likelihood estimator to accomplish this task.

Prior work [1, 4] has improved resolution uniformity. However, when one investigates the local impulse responses that quantify the local resolution properties, one finds that some asymmetries persist. Simple attempts at improving uniformity by extending the penalty neighborhood size have provided only marginal improvements. Therefore, we have sought other ways to increase design freedom.

[1] One exception to this rule is post-smoothed maximum-likelihood reconstruction. However, while the resolution properties are uniform, it generally takes much longer for an iterative algorithm to converge to a solution for an unregularized problem versus a regularized version. Additionally, preliminary investigations suggest that penalized-likelihood techniques will outperform post-smoothed maximum-likelihood in terms of noise.

II. Constrained Penalty Design

Penalized-likelihood methods have been used extensively in image reconstruction for a variety of goals. Roughness penalties promote smooth images and reduce noise in reconstructions, edge-preserving penalties can be used to include boundary information, other penalties are used to control resolution properties or optimize local contrast [5]. Regardless of the intended goal of the penalty, certain restrictions must be placed on the penalty.

Consider the penalized-likelihood objective,

$$\hat{\underline{\lambda}} = \arg\max L(\underline{\lambda}, \underline{Y}) - R(\underline{\lambda}), \tag{1}$$

where $\underline{\lambda}$ is a vector of lexicographically ordered pixel values, $\underline{Y}$ is vector measured values, $L(\underline{\lambda}, \underline{Y})$ is the likelihood term (which we will assume is convex), and $R(\underline{\lambda})$ is the penalty term. Clearly one should select $R(\underline{\lambda})$ so that finite solutions exist for (1). Additionally, one would prefer an objective that yields a unique solution. Thus, one typically uses the sufficient condition[2] of restricting $R(\underline{\lambda})$ to be a convex function.

We will focus on quadratic pairwise penalties of the form,

$$R(\underline{\lambda}) = \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} w_{jk}(\lambda_j - \lambda_k)^2 = \frac{1}{2}\underline{\lambda}^T \mathbf{R}\underline{\lambda} \tag{2}$$

$$\mathbf{R}_{jk} = \begin{cases} \sum_{l=1}^{p} \frac{1}{2}(w_{lj} + w_{jl}), & k = j \\ -w_{jk}, & k \neq j. \end{cases} , \tag{3}$$

where p is the number of pixels and w_{jk} represents a weighting that reflects the relative strength of the penalty for a given pixel pair. Traditionally, a convex $R(\underline{\lambda})$ is ensured by restricting $w_{jk} \geq 0$ for all j, k pairs. We will refer to this scheme as the *individual nonnegativity* constraints, since each interpixel weighting must be nonnegative.

The above penalty may also be written in terms of a $p \times p$ penalty matrix $\mathbf{R}$. The convexity of $R(\underline{\lambda})$ can be expressed as a nonnegative definiteness constraint on $\mathbf{R}$, or equivalently as a nonnegativity constraint on the eigenvalues of $\mathbf{R}$. If one has a cost function, $f(\mathbf{R})$, for the design of this penalty matrix, we may write

$$\hat{\mathbf{R}} = \arg \min_{\mathrm{eig}(\mathbf{R}) \geq \underline{0}} f(\mathbf{R}). \tag{4}$$

The $f(\mathbf{R})$ term could represent, for example, a least-squares design objective for the penalty, as discussed in [4], where the goal is to find a penalty that yields uniform resolution properties

[2] Alternately, it would be possible to obtain a necessary condition that depends on the form of the likelihood term, $L(\underline{\lambda}, \underline{Y})$.

in the reconstructed image. The eigenvalue constraint does not preclude negative pairwise weights between pixels.

The main problem with the formulation presented in (4) is that the minimization will generally be impractical due to the size of $\mathbf{R}$. Evaluation of the constraint, $\mathrm{eig}(\mathbf{R})$, and possibly the cost, $f(\mathbf{R})$, will generally be too computationally intensive for most applications.

In the case of a shift-invariant penalty, $\mathbf{R}$ is circulant and its eigenvalues may be computed using fast Fourier transforms. It is straightforward to formulate a shift-invariant "toy" design problem where such *Fourier constraints* may be applied. We have performed such a shift-invariant design using the linearized uniform resolution design objective presented in [4]. Specifically, we have chosen a shift-invariant problem that is similar to a design problem in the shift-variant case, where resolution anisotropy effects persist. Contours of the resulting impulse response functions are shown in Figure 1. Two things are immediately evident in this figure: (1) the Fourier constraints lead to more uniform responses, and (2) increasing the size of the penalty neighborhood does not improve uniformity for the individually applied nonnegativity constraints. Clearly the incorporation of negative weights has allowed for greater design freedom.

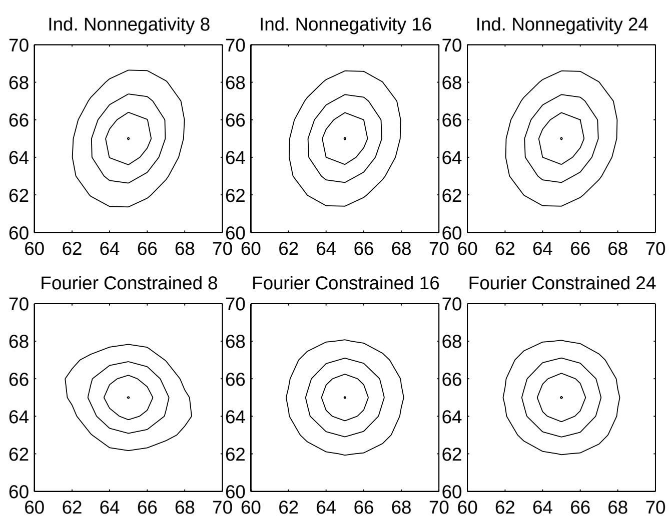
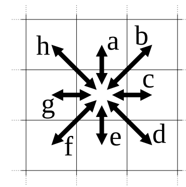


Fig. 1. Contours of the resulting impulse response functions for a toy shift-invariant problem. Two constraint methods are applied: (top row) Simple individual nonnegativity constraints and (bottom row) the full eigenvalue constraints on $\mathbf{R}$ which is applied via Fourier methods. Each column represents a design with a different penalty neighborhood size. Specifically, (left to right) results for designs with penalties including 8, 16, and 24 neighbors are shown.

The Fourier constraints are just one way of constraining a shift-invariant $\mathbf{R}$. An alternative is to use simpler constraints such as those derived by Lakshmanan for 2D Gaussian Markov random fields [6]. Unfortunately, it is unclear how to extend either these constraints or the Fourier constraints to the shift-variant case, which is required for the uniformity problem.

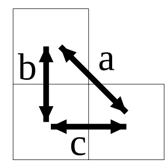
As in [4], one typically wants to develop a shift-variant penalty by performing a local design. That is, one would like to determine the weights in a pixel-by-pixel fashion, rather than all weights simultaneously. For example, at a given pixel one would like to determine all the weights between that pixel and its neighbors (see Figure 2). Unfortunately, the only pointwise constraint is the individual nonnegativity constraint. Thus, one needs to consider at least groups of pixels to incorporate negative weights.



$$\begin{array}{lll} a \geq 0 & b \geq 0 & c \geq 0 \\ d \geq 0 & e \geq 0 & f \geq 0 \\ g \geq 0 & h \geq 0 & \end{array}$$

Fig. 2. Pointwise constraints for a single pixel with eight neighbors/interpixel weights (labeled a-h).

Consider the small three pixel image shown below in Figure 3. There are three weights associated with the three pixel



$$\mathbf{R} = \begin{bmatrix} a+b & -a & -b \\ -a & a+c & -c \\ -b & -c & b+c \end{bmatrix}$$

Fig. 3. A three pixel image and its penalty matrix.

pairs, labeled a, b, and c. The penalty matrix for this image is also shown. Finding the characteristic polynomial for $\mathbf{R}$ and finding the associated Routh array, it is straightforward to derive the following constraints on the weights themselves,

$$\begin{aligned} a+b+c &\geq 0 \\ ab+bc+ac &\geq 0. \end{aligned} \tag{5}$$

These constraints allow for at most one of the weights to be negative.

While one would rarely deal with an image this small, the constraints found here can still be quite useful. Since the sum of nonnegative definite functions is nonnegative definite, one can break the summation in (2) into more manageable portions and satisfy nonnegative definiteness constraints locally. Specifically, using the constraints in (5), one can satisfy a nonnegative definiteness constraint on any sum of three weights in a large image, provided they form a loop.

A sample application of the constraints in (5) applied to a larger image is shown in Figure 4. All of the weights represented by white arrows form loops of three weights and must satisfy the constraints in (5). The remaining weights (black arrows) are not part of a loop constraint and thus must satisfy the usual individual nonnegativity constraint. Thus, the nonnegative definiteness of $\mathbf{R}$ can be guaranteed, the weights are locally constrained (allowing some form of local design), and negative weights are allowed.

These constraint loops may be chosen somewhat arbitrarily, as long as each weight is constrained exactly once (using either (5) or the simple individual nonnegativity constraint). Clearly,

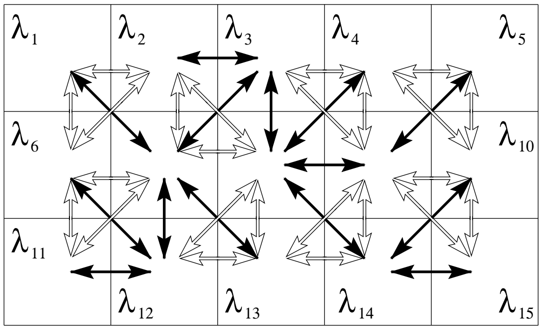


Fig. 4. A sample image, where the nonnegative definiteness of penalty has been ensured by applying the constraints in (5) over loops of three weights.

the number of ways to choose these loops increases tremendously with the number of pixels. The (impractical) optimal solution is to optimize over all possible loop configurations and select the set that yields the best $\mathbf{R}$ according to the penalty design objective, $f(\mathbf{R})$.

The best way to choose these loops will be dependent on the specific penalty design goal. In the context of the uniform resolution goal of [4], the penalty design objective is a least-squares objective that can be applied locally to find all weights linked with a given pixel position. Specifically, the objective has the form

$$\underline{\hat{w}}^j = \arg\min_{\underline{w}^j} ||\mathbf{\Phi}^j \underline{w}^j - \underline{d}^j||^2, \quad (6)$$

where $\underline{w}^j$ is a vector of weights associated with pixel j (*i.e.*, $\underline{w}^j = [w_{j1}, \ldots w_{jp}, w_{1j}, \ldots, w_{pj}]$.), and $\mathbf{\Phi}^j$ and $\underline{d}^j$ are model specific components (also dependent on pixel position). One could create an objective of the form in (4) by combining (6) for each pixel position.

To choose loops we have adopted the following heuristics:

- Calculate the unconstrained local solution, $\underline{w}^j_{\mathrm{uc}}$, to (6) for each pixel j.
- Choose only from loops that include the most negative element of $\underline{w}^j_{\mathrm{uc}}$
- Select from remaining loops by finding the loop that allows for the most negative weight. (Plug in the unconstrained solutions for the two positive values in (5) and find the bound on the remaining weight.)

While these heuristics do not necessarily yield an optimal choice for the weight constraints, such choices should generally increase design flexibility and allow for the most important negative (*i.e.*: the most negative weight in the unconstrained problem) to go negative in the constrained problem.

Once a set of constraints has been chosen (*i.e.*, a "map" such as Figure 4 is available), one must still find $\mathbf{R}$. Technically, this involves performing the minimization in (6), using the constraints that apply to these local weights, simultaneously for all pixel positions. We have opted to use an update approach where all nonlocal weights are held constant, the constrained (6) is minimized using a sequential quadratic programming algorithm, and the local solution is used as an update to the current estimate of $\mathbf{R}$. This approach is illustrated below in Figure 5. We cycle through all pixel positions until the weightings appear to have sufficiently converged.

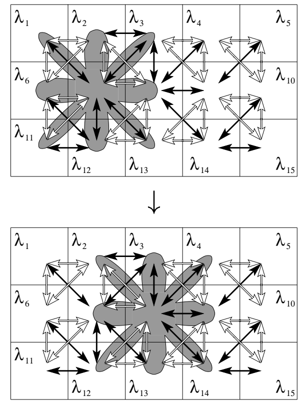


Fig. 5. An illustration of the update approach. In the first image a pointwise design is used to update only the interpixel weights lying in the gray region. Nearby weights (not in the pointwise design) that used to constrain the design are held constant. In the following image, the pointwise design is applied to the next pixel in the sequence, cycling through all pixel positions.

RESULTS

We have evaluated the these methods using the digital phantom shown in Figure 6. The measurement model includes nonuniform attenuation effects and a PET (strip integral) system model.

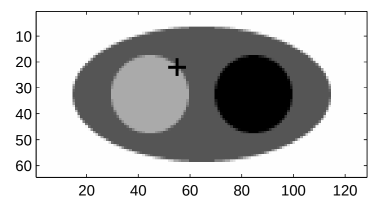


Fig. 6. Image of the emission densities for the digital phantom used in this investigation. The '+' sign indicates the position of the local impulse response used to compare the different design constraints.

We have performed the penalty design represented by (6) using both the traditional nonnegativity constraints and the proposed relaxed constraints. This penalty design used a neighborhood of the twenty nearest pixels. The designed values for one of these weighting directions (namely, the horizontal weight between neighboring pixels) is shown for the two methods below in Figure 7.

The upper half of each image shows the positive weight values and the lower half shows the negative values. For the individual nonnegativity method, there are no negative values and the lower half is blank. Also evident in the upper half of the left image is that the nonnegativity constraint is quite active for this weight. All those positions colored white indicate a zero weight. In comparison, the design with the proposed relaxed constraints does include negative weights. Additionally, if one visually combines the top and bottom halves of the right image, there are relatively few positions that are zero (*i.e.*neither positive or negative). Other weighting "directions" show similar results.

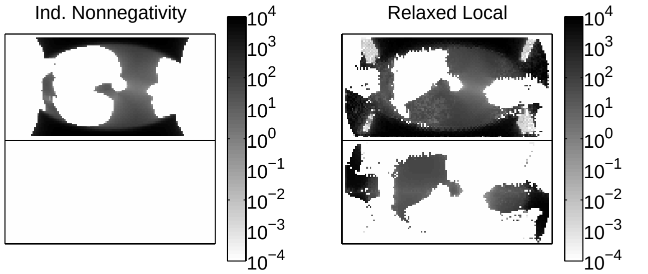


Fig. 7. Images of one of the interpixel weights (in this case, the horizontal weights) using the individual nonnegativity constraint and the relaxed local constraints. The upper half of each image shows the positive weights and the lower half shows the negative weights. Both halves of each image show weights on a logarithmic color scale.

After designing quadratic penalties under the two different constraint choices, we found the resulting local impulse responses, which quantify the resolution properties of the estimators, according to the methods in [4]. We have chosen a particular location for the local impulse response investigation. This position, indicated in Figure 6, lies in one of the areas where resolution anisotropy persisted even after the application of a penalty designed using the individual nonnegativity constraint. Local impulse responses for both the old and new constraint choices are shown in Figure 8.

Note that the local impulse response using the proposed relaxed constraints shows contours (particularly the innermost contour) closer to the desired response for which the penalty was designed (as indicated by the dashed contours). While the left image in Figure 8 shows increased blur in an slightly off-vertical direction, the right image shows improved isotropy of the response. Thus the increased design flexibility of the proposed constraints yields improved resolution uniformity.

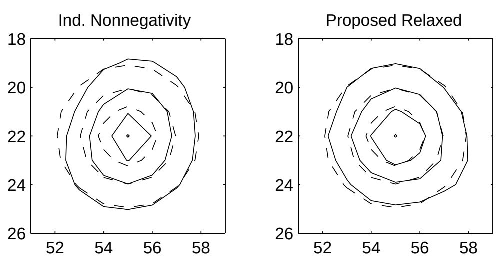


Fig. 8. Comparison of the resulting resolution properties for a penalized-likelihood PET reconstruction using the traditional nonnegatively constrained design and the proposed relaxed constraints. Contours of the desired local impulse response (dashed) and actual local impulse response (solid) are shown.

Most local impulses throughout the image show either similar improvements or or performance as good as the individual nonnegativity constraints. Indeed, we expect that the proposed constraints should yield a design no worse than the nonnegatively constrained design, since the nonnegative solution is in the feasible region of the proposed constraints. However, there are a few locations where there are slight degradations in the uniformity. We suspect that such degradations are the result of the suboptimal greedy iterative optimization approach used to calculate the weights. Such results may be due to incomplete convergence of the design optimization, or limit cycles. It is possible that some kind of regularization (*i.e.*: smoothing of the weight values between iterations) might decrease these effects. It may also be possible to develop an alternative optimization approach that is less susceptible to such problems. However, it should also be noted that these effects are generally relatively small, and that the uniformity improvements outweigh the degradations.

III. Conclusions

We have found that the nonnegativity constraints in our prior penalty design work are often very active, and that a zero weighting still appears to induce too much smoothing in certain directions, yielding anisotropic resolution properties. The ability to include negative weights allows one to more effectively reorient the penalty “direction.” While shifting some weights to be negative will generally increase the values of other weights to satisfy (5), the nonnegatively constrained feasible region is a subset of the feasible region imposed by the proposed constraints, and thus (theoretically) yields solutions no worse than those found using the traditional constraints. While it is difficult to find the optimal solution in practice, we have demonstrated a simple practical suboptimal approach that can be implemented to obtain improved uniformity performance.

The proposed constraints can be applied locally, are computationally practical, and yield improved performance for uniform resolution penalty design. While these constraints were developed for the goal of increased flexibility in resolution control, these techniques may be applicable in other situations. That is, there may be other penalty objectives where a similar increase in design flexibility may be important.

It is possible that one could find other constraints using a similar development. That is, instead of using a small region of three pixels as in (3), one could extend these methods to incorporate weight constraints on larger loops or other weight geometries. However, while making the local constraints increasingly complicated may increase local design flexibility, it will become increasingly difficult to incorporated the local constraints into a global set of constraints. In other words, it becomes increasingly difficult to form the analogous constraint map to the one shown in Figure 4, both in terms of the actual geometric fitting of such constraints and the selection of which constraints will yield the greatest (or most significant) design freedom.

References

[1] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs,” *IEEE Tr. Im. Proc.*, vol. 5, pp. 1346–58, Sept. 1996.

[2] W. Xia, R. M. Lewitt, and P. R. Edholm, “Fourier correction for spatially variant collimator blurring in SPECT,” *IEEE Tr. Med. Im.*, vol. 14, pp. 100–15, Mar. 1995.

[3] E. J. Soares, C. L. Byrne, S. J. Glick, C. R. Appledorn, and M. A. King, “Implementation and evaluation of and analytical solution to the photon attenuation and nonstationary resolution reconstruction problem in SPECT,” *IEEE Tr. Nuc. Sci.*, vol. 40, pp. 1231–1237, Aug. 1993.

[4] J. W. Stayman and J. A. Fessler, “Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction,” *IEEE Tr. Med. Im.*, vol. 19, pp. 601–15, June 2000.

[5] J. Qi and R. M. Leahy, “A theoretical study of the contrast recovery and variance of MAP reconstructions with applications to the selection of

smoothing parameters," *IEEE Tr. Med. Im.*, vol. 18, pp. 293–305, Apr. 1999.

[6] S. Lakshmanan and H. Derin, "Valid parameter space for 2-D Gaussian Markov random fields," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 39, pp. 703–9, Mar. 1993.