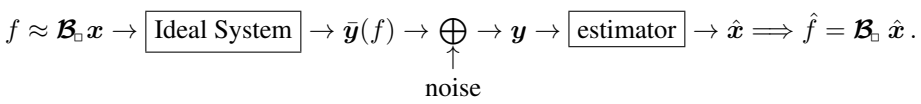
Chapter 22

Spatial Resolution Properties

ch,srp

Contents

22.1 Introduction (s,srp,intro) . . . **22.2**
22.2 Definitions of impulse response (s,srp,def) . . . **22.2**
22.2.1 Discrete-discrete local impulse response . . . 22.2
22.2.1.1 Data-dependent definition . . . 22.2
22.2.1.2 Data mean definition . . . 22.2
22.2.1.3 Estimator mean definition . . . 22.3
22.2.2 Continuous-discrete local impulse response . . . 22.3
22.3 Estimator gradient (s,srp,grad) . . . **22.4**
22.4 Penalized-likelihood estimators (s,srp,pl) . . . **22.5**
22.4.1 Independent measurements (s,srp,pl,ind) . . . 22.6
22.4.2 Image reconstruction case . . . 22.7
22.4.3 Examples . . . 22.7
22.4.3.1 Penalized weighted least squares (PWLS) . . . 22.7
22.4.3.2 Quadratically penalized weighted least squares (QPWLS) . . . 22.7
22.4.3.3 Poisson emission penalized likelihood . . . 22.8
22.4.3.4 Poisson transmission penalized likelihood . . . 22.8
22.5 Computing the local impulse response (s,srp,comp) . . . **22.8**
22.5.1 Perturbation approximation . . . 22.8
22.5.2 Monte Carlo estimator . . . 22.8
22.5.3 Unbiased estimator . . . 22.9
22.5.4 Matrix-based approximations . . . 22.9
22.6 Local shift-invariance (s,srp,local) . . . **22.9**
22.7 Local Fourier approximations (s,srp,dft) . . . **22.10**
22.8 Complex data, real images (s,srp,complex) . . . **22.11**
22.9 Fisher information approximations (s,srp,fish) . . . **22.11**
22.9.1 Certainty approximation . . . 22.11
22.9.2 Angular dependent approximation . . . 22.12
22.10 Regularization design based on spatial resolution properties (s,srp,design) . . . **22.13**
22.10.1 Certainty-based design . . . 22.13
22.10.2 Analytical Fourier-based design . . . 22.14
22.11 Resolution analysis of iterations (s,srp,iter) . . . **22.14**
22.12 PSF of denoising with quadratic regularization (s,srp,denoise) . . . **22.15**
22.12.1 1D analysis of impulse response (s,srp,dtft) . . . 22.15
22.12.2 N-D analysis . . . 22.16
22.13 Dynamic problems (s,srp,dynam) . . . **22.17**
22.13.1 Static system matrix . . . 22.18
22.13.1.1 Static object case . . . 22.18
22.13.2 General system matrices . . . 22.18
22.13.2.1 Static object case . . . 22.19
22.14 Notes (s,srp,note) . . . **22.20**
22.15 Problems (s,srp,prob) . . . **22.21**
22.16 Bibliography . . . **22.21**



fig,srp,system

Figure 22.1.1: General model for imaging systems and reconstruction methods.

22.1 Introduction (s,srp,intro)

s,srp,intro

This chapter analyzes the *spatial resolution properties* of image reconstruction methods. This analysis is important for understanding the trade-offs between spatial resolution and noise, both for imaging system design and for fine tuning the estimators. In particular, analysis of resolution properties is important for the design of regularization methods that lead to reconstructed images with desired properties, such as uniform spatial resolution.

The definition of "resolution" in imaging has proven to be a somewhat elusive concept, dating back at least to Rayleigh's criterion [1]. Numerous papers have considered various definitions and surveys thereof *e.g.*, [2–5], particularly in the context of *super resolution* in *diffraction limited* optical imaging *e.g.*, [6–11]

We analyze spatial resolution in the context of the block diagram shown in Fig. 22.1.1. An image reconstruction method operates on a measurement vector $\boldsymbol{y} \in \mathbb{C}^{n_\mathrm{d}}$ to yield an estimated object vector $\hat{\boldsymbol{x}}(\boldsymbol{y}) \in \mathbb{C}^{n_\mathrm{p}}$. If the estimator were a linear function of $\boldsymbol{y}$, and if $\bar{\boldsymbol{y}} = \mathsf{E}[\boldsymbol{y}]$ were a linear function of the true object f^{true}, then analysis of the *spatial resolution properties* of the estimator would be relatively straightforward. However, many of the image reconstruction methods described in this book are nonlinear, so spatial resolution analysis is more subtle. Furthermore, even for linear image reconstruction methods, spatial resolution analysis can be complicated by shift-varying system or estimator properties.

For nonlinear estimators, the local resolution properties can depend on many factors: the imaging system's resolution properties, the measurement statistics, the type of regularization used, and even the unknown object f^{true} itself. To analyze the resolution properties in such situations, we examine the *local impulse response*. Intuitively, the local impulse response should describe how the estimate $\hat{\boldsymbol{x}}$ would change due to a point-like perturbation of the true object at a given spatial location. In the case of nonlinear estimators, there is not a single canonical way to define spatial resolution. The next section describes a few different ways to formalize this concept mathematically.

22.2 Definitions of impulse response (s,srp,def)

s,srp,def

There are two families of definitions for *local impulse response* functions, depending on whether one considers the entire problem as a discrete-space problem, or whether one acknowledges that the underlying object is continuous (even if the reconstruction is based on a vector estimate $\hat{\boldsymbol{x}}$). This section considers both frameworks.

22.2.1 Discrete-discrete local impulse response

We focus first on the discrete-space formulation, *i.e.*, a discrete-space object $\boldsymbol{x}_{\mathrm{true}}$ leads to a noisy measurement vector $\boldsymbol{y}$ with mean $\mathsf{E}[\boldsymbol{y}] = \bar{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{true}})$ from which we compute a discrete-space estimate $\hat{\boldsymbol{x}}(\boldsymbol{y})$. Each definition below provides a different way of quantifying how a perturbation of the object of the form $\boldsymbol{x} + \varepsilon \boldsymbol{e}_j$ will influence the estimator $\hat{\boldsymbol{x}}$, where $\boldsymbol{e}_j$ denotes the jth unit vector in $\mathbb{R}^{n_\mathrm{p}}$.

22.2.1.1 Data-dependent definition

The following definition of *local impulse response* depends on the particular noisy measurement vector $\boldsymbol{y}$:

$$\begin{aligned} \bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) &\triangleq \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{x}}(\boldsymbol{y} + [\bar{\boldsymbol{y}}(\boldsymbol{x} + \varepsilon \boldsymbol{e}_j) - \bar{\boldsymbol{y}}(\boldsymbol{x})]) - \hat{\boldsymbol{x}}(\boldsymbol{y})}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{x}}(\boldsymbol{y} + \varepsilon \nabla \bar{\boldsymbol{y}}(\boldsymbol{x})\, \boldsymbol{e}_j) - \hat{\boldsymbol{x}}(\boldsymbol{y})}{\varepsilon} = \nabla \hat{\boldsymbol{x}}(\boldsymbol{y})\ \nabla \bar{\boldsymbol{y}}(\boldsymbol{x})\ \boldsymbol{e}_j, \end{aligned} \tag{22.2.1}$$

where $\nabla \hat{\boldsymbol{x}}$ is a $n_\mathrm{p} \times n_\mathrm{d}$ matrix and $\nabla \bar{\boldsymbol{y}}$ and is a $n_\mathrm{d} \times n_\mathrm{p}$ matrix. The quantity $\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x})$ describes how much the estimator $\hat{\boldsymbol{x}}$ changes due to a perturbation of the jth pixel in the true object $\boldsymbol{x}$. Often $\bar{\boldsymbol{y}}$ is affine in $\boldsymbol{x}$, *i.e.*, $\bar{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}$, in which case $\nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{A}$. If $\hat{\boldsymbol{x}}$ were an affine function of $\boldsymbol{y}$, *e.g.*, $\hat{\boldsymbol{x}} = \boldsymbol{Z}\boldsymbol{y}$ then this definition would simplify to that of the usual (local) impulse response: $\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) = \boldsymbol{Z}\boldsymbol{A}\boldsymbol{e}_j$.

22.2.1.2 Data mean definition

Often we would like to examine separately the spatial resolution properties and the noise properties of a reconstruction method, so the inclusion of a noisy vector $\boldsymbol{y}$ in the definition (22.2.1) can be unnatural. It may seem more reasonable

to define the local impulse response in terms of the ensemble *mean* of the measurements:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) \triangleq \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x} + \varepsilon \boldsymbol{e}_j)) - \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x}))}{\varepsilon} = \nabla \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x})) \, \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \, \boldsymbol{e}_j. \tag{22.2.2}$$

e,srp,lir,ljx

This definition results in the same expression as (22.2.1) but with $\boldsymbol{y}$ replaced by $\bar{\boldsymbol{y}}$. In practice, we usually use $\bar{\boldsymbol{y}}$ for simulation studies, but for real data we usually use $\boldsymbol{y}$, because $\bar{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{true}})$ is unavailable for real data. Alternatively, the "plug in" estimate $\bar{\boldsymbol{y}}(\hat{\boldsymbol{x}})$ is often an adequate approximation to $\bar{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{true}})$ for the purposes of resolution analysis.

22.2.1.3 Estimator mean definition

Another natural definition is to express the resolution properties in terms of the *mean* reconstructed object vector:

$$\boldsymbol{\mu}(\boldsymbol{x}) \triangleq \mathsf{E}_{\boldsymbol{x}}[\hat{\boldsymbol{x}}(\boldsymbol{y})] = \int \hat{\boldsymbol{x}}(\boldsymbol{y}) \, \mathsf{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \, \mathrm{d}\boldsymbol{y}, \tag{22.2.3}$$

e,srp,mean,x

leading to the following definition:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) \triangleq \lim_{\varepsilon \to 0} \frac{\boldsymbol{\mu}(\boldsymbol{x} + \varepsilon \boldsymbol{e}_j) - \boldsymbol{\mu}(\boldsymbol{x})}{\varepsilon} = \nabla \boldsymbol{\mu}(\boldsymbol{x}) \, \boldsymbol{e}_j. \tag{22.2.4}$$

e,srp,lir,mu,x

For general nonlinear estimators it is very difficult to determine the mean function $\boldsymbol{\mu}(\cdot)$ exactly. In practice, we often assume that the mean $\boldsymbol{\mu}$ approximately equals the estimate one would obtain from noiseless data:

$$\boldsymbol{\mu}(\boldsymbol{x}) = \mathsf{E}_{\boldsymbol{x}}[\hat{\boldsymbol{x}}(\boldsymbol{y})] \approx \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x})) \,. \tag{22.2.5}$$

e,srp,mean,approx,x

If this approximation is accurate, then $\nabla \boldsymbol{\mu}(\boldsymbol{x}) = \nabla \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x})$, so the definitions (22.2.4) and (22.2.2) become equivalent. Therefore, in the usual case where an exact expression for $\boldsymbol{\mu}(\boldsymbol{x})$ is unavailable, it seems more "honest" to use (22.2.2) as the definition of local impulse response directly rather than using (22.2.4) and then invoking the approximation (22.2.5). In fact, the definition (22.2.1) seems the most general because one can always substitute $\bar{\boldsymbol{y}}$ for $\boldsymbol{y}$ in (22.2.1) if desired.

Each of the above definitions involves the measurement mean vector $\bar{\boldsymbol{y}}$ or its gradient. So one must have a system model $\bar{\boldsymbol{y}}(\boldsymbol{x})$ to analyze spatial resolution properties. If that system model is inaccurate, then the analytical predictions may poorly match the resolution properties of the estimator when applied to real data.

s,srp,def,cont

22.2.2 Continuous-discrete local impulse response

In practice, the true object f^{true} is a continuous-space function, even though we compute a finite-dimensional estimate $\hat{\boldsymbol{x}}$. (Subsequently we can form an estimate $\hat{f}$ using the object basis function expansion (10.1.1).) As an alternative method for quantifying resolution properties, one could explore how an impulse at some continuous-space location $\vec{\mathrm{x}}$ affects the estimator $\hat{\boldsymbol{x}}$.

One definition of local impulse response is specific to each noisy measurement vector $\boldsymbol{y}$:

$$\bar{\boldsymbol{l}}(\boldsymbol{y}, \vec{\mathrm{x}}) \triangleq \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{x}}(\boldsymbol{y} + \varepsilon \boldsymbol{p}(\vec{\mathrm{x}})) - \hat{\boldsymbol{x}}(\boldsymbol{y})}{\varepsilon} = \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) \, \boldsymbol{p}(\vec{\mathrm{x}}), \tag{22.2.6}$$

e,srp,lir,lye

where the differential change in the mean measurements due to a perturbation of the true object at spatial location $\vec{\mathrm{x}}$ is denoted

$$\boldsymbol{p}(\vec{\mathrm{x}}) \triangleq \Delta \bar{\boldsymbol{y}}(f, \vec{\mathrm{x}}) \,.$$

In particular, for an affine model of the form $\bar{y}_i = \int \mathsf{s}_i(\vec{\mathrm{x}}) \, f(\vec{\mathrm{x}}) \, \mathrm{d}\vec{\mathrm{x}} + r_i$, then the perturbation vector $\boldsymbol{p}(\vec{\mathrm{x}})$ has elements $p_i(\vec{\mathrm{x}}) = \mathsf{s}_i(\vec{\mathrm{x}})$, $i = 1, \ldots, n_{\mathrm{d}}$.

Alternatively it may seem more reasonable to define the local impulse response in terms of the *mean* measurements:

$$\bar{\boldsymbol{l}}(f, \vec{\mathrm{x}}) \triangleq \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(f + \varepsilon \delta_{\vec{\mathrm{x}}})) - \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(f))}{\varepsilon} = \nabla \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(f)) \, \boldsymbol{p}(\vec{\mathrm{x}}), \tag{22.2.7}$$

e,srp,lir,yb

where $\delta_{\vec{\mathrm{x}}}$ denotes a *Dirac impulse* centered at spatial location $\vec{\mathrm{x}}$, and where $\boldsymbol{p}$ was defined above. This definition results in the same expression as (22.2.6) but with $\boldsymbol{y}$ replaced by $\bar{\boldsymbol{y}}$.

Another natural definition is to express the resolution properties in terms of the *mean* reconstructed object vector:

$$\boldsymbol{\mu}(f) \triangleq \mathsf{E}_f[\hat{\boldsymbol{x}}(\boldsymbol{y})] = \int \hat{\boldsymbol{x}}(\boldsymbol{y}) \, \mathsf{p}(\boldsymbol{y} \,|\, f) \, \mathrm{d}\boldsymbol{y}, \tag{22.2.8}$$

e,srp,mean

leading to

$$\bar{\boldsymbol{l}}(f, \vec{\mathrm{x}}) \triangleq \lim_{\varepsilon \to 0} \frac{\boldsymbol{\mu}(f + \varepsilon \delta_{\vec{\mathrm{x}}}) - \boldsymbol{\mu}(f)}{\varepsilon} = \Delta \boldsymbol{\mu}(f, \vec{\mathrm{x}}) \,. \tag{22.2.9}$$

e,srp,lir,mu

Again, for general nonlinear estimators it is very difficult to determine the mean $\boldsymbol{\mu}$ exactly. One could invoke the following approximation:

e,srp,mean,approx

$$\boldsymbol{\mu}(f) = \mathsf{E}_f[\hat{\boldsymbol{x}}(\boldsymbol{y})] \approx \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(f)), \tag{22.2.10}$$

in which case the definitions (22.2.9) and (22.2.7) are equivalent. Or one can simply use (22.2.7) as the definition of local impulse response rather than using (22.2.9) and then invoking the approximation (22.2.10).

Each of the above definitions involves the measurement mean vector $\bar{\boldsymbol{y}}$. So one must have a continuous-discrete system model $\bar{\boldsymbol{y}}(f)$ to analyze spatial resolution properties using any of these definitions.

22.3 Estimator gradient (s,srp,grad)

s,srp,grad

All of the definitions of local impulse response given in §22.2 lead to expressions that depend on the gradient of the estimator $\hat{\boldsymbol{x}}(\cdot)$. Thus to determine the local impulse response the key step is to find the estimator gradient $\nabla \hat{\boldsymbol{x}}(\boldsymbol{y})$. If $\hat{\boldsymbol{x}}$ is affine in $\boldsymbol{y}$, *i.e.*, if $\hat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{u}$ for some $n_\mathrm{p} \times n_\mathrm{d}$ matrix $\boldsymbol{A}$, then of course the gradient is simply $\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{A}$. But many estimators of interest are defined *implicitly* as the minimizer of a cost function:

e,srp,xh

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}, \boldsymbol{y}). \tag{22.3.1}$$

To analyze the gradient of $\hat{\boldsymbol{x}}$ in this case, we assume that the cost function Ψ satisfies the following regularity conditions.

- For each $\boldsymbol{y}$, there exists a unique minimizer of $\Psi(\cdot, \boldsymbol{y})$.
- Ψ is differentiable with respect to $\boldsymbol{x}$, *i.e.*, $\nabla^{[1,0]} \Psi$ is well defined, where $\nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ denotes the $n_\mathrm{p} \times 1$ column vector with elements

e,srp,grad,10

$$\left[\nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})\right]_j = \frac{\partial}{\partial x_j} \Psi(\boldsymbol{x}, \boldsymbol{y}). \tag{22.3.2}$$

- $\nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ is continuously differentiable with respect to both its arguments, *i.e.*, $\nabla^{[2,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ denotes the $n_\mathrm{p} \times n_\mathrm{p}$ Hessian matrix with elements

$$\left[\nabla^{[2,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})\right]_{jk} = \frac{\partial^2}{\partial x_j \, \partial x_k} \Psi(\boldsymbol{x}, \boldsymbol{y}),$$

  and $\nabla^{[1,1]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ denotes the $n_\mathrm{p} \times n_\mathrm{d}$ matrix with elements

e,srp,grad,11

$$\left[\nabla^{[1,1]} \Psi(\boldsymbol{x}, \boldsymbol{y})\right]_{ji} = \frac{\partial^2}{\partial x_j \, \partial y_i} \Psi(\boldsymbol{x}, \boldsymbol{y}). \tag{22.3.3}$$

- The Hessian $\nabla^{[2,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ is invertible. (A sufficient condition for this would be for Ψ to be strictly convex.)

Then by the *implicit function theorem* [12, p. 331], the estimator $\hat{\boldsymbol{x}}(\boldsymbol{y})$ is a well-defined and continuously differentiable function of $\boldsymbol{y}$.

Disregarding any constraints (such as the nonnegativity), the minimizer must satisfy

e,srp,grad=0

$$\boldsymbol{0}_{n_\mathrm{p} \times 1} = \nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \Big|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})} = \nabla^{[1,0]} \Psi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}), \tag{22.3.4}$$

where $\boldsymbol{0}_{n \times m}$ denotes the $n \times m$ matrix of zeros. Next we apply the chain rule to differentiate (22.3.4) with respect to $\boldsymbol{y}$ as follows:

e,srp,grad,chain

$$\boldsymbol{0}_{n_\mathrm{p} \times n_\mathrm{d}} = \nabla^{[2,0]} \Psi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) + \nabla^{[1,1]} \Psi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}). \tag{22.3.5}$$

Rearranging and solving yields the following general expression for the estimator gradient:

e,srp,grad,xh

$$\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = \left[\nabla^{[2,0]} \Psi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y})\right]^{-1} \left[-\nabla^{[1,1]} \Psi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y})\right]. \tag{22.3.6}$$

This key expression is the foundation for all subsequent analysis.

x,srp,wls

Example 22.3.1 *Consider the WLS estimator with cost function* $\Psi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2_{\boldsymbol{W}^{1/2}}$. *Then one can show that* $\nabla^{[1,0]} \Psi = -\boldsymbol{A}'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$, $\nabla^{[2,0]} \Psi = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$, *and* $-\nabla^{[1,1]} \Psi = \boldsymbol{A}'\boldsymbol{W}$. *So the estimator gradient is* $\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}]^{-1} \boldsymbol{A}'\boldsymbol{W}$. *This expression is consistent with the fact that this estimator is linear,* i.e., $\hat{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{Z}\boldsymbol{y}$. *where* $\boldsymbol{Z} = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}]^{-1} \boldsymbol{A}'\boldsymbol{W}$.

Note that we disregarded the nonnegativity constraint in (22.3.1). In problems with such constraints, one should consider the estimator gradient expression (22.3.6) to be an approximation, the accuracy of which can be suspect in image regions where the nonnegativity constraint is active.

Distribution	$g(z,y)$	$g^{[1,0]}(z,y)$	$g^{[2,0]}(z,y)$	$-g^{[1,1]}(z,y)$
Normal	$\frac{1}{2\sigma^2}(z-y)^2$	$\frac{1}{\sigma^2}(z-y)$	$\frac{1}{\sigma^2}$	$\frac{1}{\sigma^2}$
Symmetric: $\psi(t) = \psi(-t)$	$\psi(\ell - y)$	$\dot{\psi}(\ell - y)$	$\ddot{\psi}(\ell - y)$	$\ddot{\psi}(\ell - y)$
Poisson	$z - y \log z$	$1 - \frac{y}{z}$	$\frac{y}{z^2}$	$\frac{1}{z}$

Table 22.1: Derivatives of marginal negative log-likelihoods $g(z,y)$ for various measurement noise models. (Irrelevant additive constants independent of z are ignored.) t,srp,table,g

22.4 Penalized-likelihood estimators (s,srp,pl)

s,srp,pl

Penalized-likelihood estimators are based on cost functions of the form

$$\Psi(\boldsymbol{x}, \boldsymbol{y}) = \text{Ł}(\boldsymbol{x}, \boldsymbol{y}) + \mathsf{R}(\boldsymbol{x}), \tag{22.4.1}$$

e,srp,pl,kost

where Ł denotes the negative log-likelihood and where the regularization function $\mathsf{R}(\boldsymbol{x})$ is usually independent[1] of the data $\boldsymbol{y}$. For such cost functions, the estimator gradient (22.3.6) becomes

$$\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = \left[\nabla^{[2,0]} \text{Ł}(\boldsymbol{x}, \boldsymbol{y}) + \mathbf{R}(\boldsymbol{x})\right]^{-1} \left[-\nabla^{[1,1]} \text{Ł}(\boldsymbol{x}, \boldsymbol{y})\right] \Big|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})}, \tag{22.4.2}$$

e,srp,pl,grad,xh

where $\mathbf{R}(\boldsymbol{x}) = \nabla^2 \mathsf{R}(\boldsymbol{x})$ denotes the Hessian of the regularizer. Combining with (22.2.1) yields the following expression for the *local impulse response*:

$$\begin{aligned} \bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) &= \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \boldsymbol{e}_j \\ &= \left[\nabla^{[2,0]} \text{Ł}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) + \mathbf{R}(\hat{\boldsymbol{x}}(\boldsymbol{y}))\right]^{-1} \left[-\nabla^{[1,1]} \text{Ł}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y})\right] \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \boldsymbol{e}_j. \end{aligned} \tag{22.4.3}$$

Typically one expects that as $\mathbf{R}$ approaches zero, the local impulse response should approach the Kronecker impulse $\boldsymbol{e}_j$, at least for reasonably well-conditioned problems. This property is not evident immediately in (22.4.3). However, for most statistical models used in imaging problems, the negative log-likelihood Ł is minimized at the true parameter $\boldsymbol{x}$ when given noiseless data. In other words, usually $\boldsymbol{x} = \arg\min_{\boldsymbol{z}} \text{Ł}(\boldsymbol{z}, \bar{\boldsymbol{y}}(\boldsymbol{x}))$ or equivalently

$$\nabla^{[1,0]} \text{Ł}(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{x})) = \mathbf{0}. \tag{22.4.4}$$

e,srp,pl,noiseless

Differentiating both sides with respect to $\boldsymbol{x}$ yields

$$\nabla^{[2,0]} \text{Ł}(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{x})) + \nabla^{[1,1]} \text{Ł}(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) = \mathbf{0}.$$

We would like to combine this with (22.4.3) but the terms do not exactly match in general. To proceed, first we adopt the definition (22.2.2) where $\boldsymbol{y} \mapsto \bar{\boldsymbol{y}}(\boldsymbol{x})$, yielding

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) = \bar{\boldsymbol{l}}^{(j)}(\bar{\boldsymbol{y}}(\boldsymbol{x}), \boldsymbol{x}) = \left[\nabla^{[2,0]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) + \mathbf{R}(\check{\boldsymbol{x}})\right]^{-1} \left[-\nabla^{[1,1]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x}))\right] \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \boldsymbol{e}_j,$$

where $\check{\boldsymbol{x}} \triangleq \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x}))$. Still the terms do not exactly match because of the $\check{\boldsymbol{x}}$ terms. However, as the regularization strength decreases, usually $\check{\boldsymbol{x}}$ approaches $\boldsymbol{x}$, so

$$-\nabla^{[1,1]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \approx \nabla^{[2,0]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})).$$

Substituting this yields the following final general expression for the local impulse response of penalized-likelihood estimators that satisfy (22.4.4):

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) \approx \left[\nabla^{[2,0]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) + \mathbf{R}(\check{\boldsymbol{x}})\right]^{-1} \nabla^{[2,0]} \text{Ł}(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \boldsymbol{e}_j. \tag{22.4.5}$$

e,srp,pl,ljyx,20

In this expression it is clear that $\bar{\boldsymbol{l}}^{(j)}$ approaches $\boldsymbol{e}_j$ as $\mathbf{R}$ approaches zero, for well-conditioned problems.

[1] Some of the regularization design methods described in §(penalty design) are data dependent. Determining the effect of this dependence on resolution properties is an *open problem.*

Distribution	$\mathsf{h}(\ell, y)$	$\mathsf{h}^{[1,0]}(\ell, y)$	$\mathsf{h}^{[2,0]}(\ell, y)$	$-\mathsf{h}^{[1,1]}(\ell, y)$
Normal	$\frac{1}{2\sigma^2}(\ell - y)^2$	$\frac{1}{\sigma^2}(\ell - y)$	$\frac{1}{\sigma^2}$	$\frac{1}{\sigma^2}$
Symmetric: $\psi(t) = \psi(-t)$	$\psi(\ell - y)$	$\dot{\psi}(\ell - y)$	$\ddot{\psi}(\ell - y)$	$\ddot{\psi}(\ell - y)$
Poisson General	$\bar{y}_i(\ell) - y \log \bar{y}_i(\ell)$	$\dot{\bar{y}}_i \left[1 - \frac{y}{\bar{y}_i}\right]$	$\ddot{\bar{y}}_i \left[1 - \frac{y}{\bar{y}_i}\right] + y \frac{\dot{\bar{y}}_i^2}{\bar{y}_i^2}$	$\frac{\dot{\bar{y}}_i}{\bar{y}_i}$
Poisson Emission	$\bar{y}_i(\ell) = c\ell + r$	$c \left[1 - \frac{y}{\bar{y}_i}\right]$	$y \frac{c^2}{\bar{y}_i^2}$	$\frac{c}{\bar{y}_i}$
Poisson Transmission	$\bar{y}_i(\ell) = b\,\mathrm{e}^{-\ell} + r$	$-b\,\mathrm{e}^{-\ell} \left[1 - \frac{y}{\bar{y}_i}\right]$	$\left[1 - \frac{yr}{\bar{y}_i^2}\right] b\,\mathrm{e}^{-\ell}$	$-\left[1 - \frac{r}{\bar{y}_i}\right]$

Table 22.2: Derivatives of marginal negative log-likelihoods $\mathsf{h}(\ell, y)$ for various measurement noise models. (Irrelevant additive constants independent of ℓ are ignored.) t,srp,table,h

s,srp,pl,ind

22.4.1 Independent measurements (s,srp,pl,ind)

In the usual case where the measurements $\{y_i\}$ are statistically independent, the negative log-likelihood has the *additively separable* form

$$\mathsf{L}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n_\mathrm{d}} g_i(\bar{y}_i(\boldsymbol{x}), y_i) \qquad (22.4.6)$$ e,srp,pl,ind,Lxy

for some functions $\{g_i(\cdot, \cdot)\}$ that depend on the statistical model. For example, for a Poisson noise model, $g_i(z, y) = z - y \log z$. Table 22.1 lists more examples. Such *additively separable* log-likelihoods have the following gradients:

$$\begin{aligned}
\left[\nabla^{[1,0]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y})\right]_j &= \sum_{i=1}^{n_\mathrm{d}} g_i^{[1,0]}(\bar{y}_i(\boldsymbol{x}), y_i) \frac{\partial}{\partial x_j} \bar{y}_i(\boldsymbol{x}) \\
\left[\nabla^{[1,1]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y})\right]_{ji} &= g_i^{[1,1]}(\bar{y}_i(\boldsymbol{x}), y_i) \frac{\partial}{\partial x_j} \bar{y}_i(\boldsymbol{x}) \\
\left[\nabla^{[2,0]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y})\right]_{jk} &= \sum_{i=1}^{n_\mathrm{d}} g_i^{[2,0]}(\bar{y}_i(\boldsymbol{x}), y_i) \frac{\partial}{\partial x_j} \bar{y}_i(\boldsymbol{x}) \frac{\partial}{\partial x_k} \bar{y}_i(\boldsymbol{x}) \\
&\quad + g_i^{[1,0]}(\bar{y}_i(\boldsymbol{x}), y_i) \frac{\partial^2}{\partial x_j \, \partial x_k} \bar{y}_i(\boldsymbol{x}),
\end{aligned} \qquad (22.4.7)$$

for $k, j = 1, \ldots, n_\mathrm{p}$ and $i = 1, \ldots, n_\mathrm{d}$.

For most statistical models, $g_i(z, y)$ is minimized (has highest log-likelihood) when $z = y$. In other words[2]:

$$g_i^{[1,0]}(y, y) = 0, \qquad (22.4.8)$$ e,srp,pl,ind,noiseless

cf. (22.4.8). Furthermore, $g_i^{[1,0]}(z, y)$ is usually smooth, so $g_i^{[1,0]}(z, y) \approx 0$ for $z \approx y$. And typically we will consider cases where $\bar{y}_i(\boldsymbol{x}) \approx y_i$. In addition, $\frac{\partial^2}{\partial x_j \, \partial x_k} \bar{y}_i(\boldsymbol{x})$ is exactly zero for many system models. Thus, we disregard the second term in (22.4.7), leading to the following matrix expressions for the gradients:

$$\begin{aligned}
-\nabla^{[1,1]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y}) &= (\nabla \bar{\boldsymbol{y}}(\boldsymbol{x}))' \boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) \\
\nabla^{[2,0]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y}) &\approx (\nabla \bar{\boldsymbol{y}}(\boldsymbol{x}))' \boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y}) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}),
\end{aligned} \qquad (22.4.9)$$

where we define the following diagonal matrices:

$$\begin{aligned}
\boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) &\triangleq \mathsf{diag}\left\{ - g_i^{[1,1]}(\bar{y}_i(\boldsymbol{x}), y_i) \right\} \\
\boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y}) &\triangleq \mathsf{diag}\left\{ g_i^{[2,0]}(\bar{y}_i(\boldsymbol{x}), y_i) \right\}.
\end{aligned} \qquad (22.4.10)$$

Combining this with (22.4.3) yields the following expression for the local impulse response of a penalized-likelihood estimator with independent measurements:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) \approx \left[(\nabla \bar{\boldsymbol{y}}(\boldsymbol{z}))' \boldsymbol{D}_2(\boldsymbol{z}, \boldsymbol{y}) \nabla \bar{\boldsymbol{y}}(\boldsymbol{z}) + \mathsf{R}(\boldsymbol{z}) \right]^{-1} (\nabla \bar{\boldsymbol{y}}(\boldsymbol{z}))' \boldsymbol{D}_1(\boldsymbol{z}, \boldsymbol{y}) \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \, \boldsymbol{e}_j \Big|_{\boldsymbol{z} = \hat{\boldsymbol{x}}(\boldsymbol{y})}. \qquad (22.4.11)$$ e,srp,pl,ind,ljyx

[2] One exception is a saddle point approximation to the log-likelihood of the *compound Poisson* model for X-ray CT statistics [13].

This expression is exact in cases where $\frac{\partial^2}{\partial x_j \partial x_k} \bar{y}_i(\boldsymbol{x})$ is zero, and in cases where $\boldsymbol{y} = \bar{\boldsymbol{y}}(\boldsymbol{x})$, and is approximate otherwise.

In the usual cases that satisfy (22.4.8), differentiating again yields

$$-g_i{}^{[1,1]}(y,y) = g_i{}^{[2,0]}(y,y).$$

Thus $\boldsymbol{D}_1(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{x})) = \boldsymbol{D}_2(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{x}))$, but this is not quite what is needed to simplify (22.4.11). However, $\check{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x}))$ approaches $\boldsymbol{x}$ as the regularization R decreases, so $\boldsymbol{D}_1(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \approx \boldsymbol{D}_2(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x}))$, leading to the final approximation for the local impulse response for independent measurements:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) \approx \left[(\nabla \bar{\boldsymbol{y}}(\check{\boldsymbol{x}}))' \boldsymbol{D}_2(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \nabla \bar{\boldsymbol{y}}(\check{\boldsymbol{x}}) + \mathsf{R}(\check{\boldsymbol{x}}) \right]^{-1} (\nabla \bar{\boldsymbol{y}}(\check{\boldsymbol{x}}))' \boldsymbol{D}_2(\check{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{x})) \nabla \bar{\boldsymbol{y}}(\check{\boldsymbol{x}}) \boldsymbol{e}_j. \tag{22.4.12}$$

s,srp,pl,ir

22.4.2 Image reconstruction case

For most image reconstruction problems, $\bar{y}_i(\boldsymbol{x})$ is a function of $[\boldsymbol{A}\boldsymbol{x}]_i$ for some system matrix $\boldsymbol{A}$, in which case the marginal log-likelihoods g_i have the following form:

srp,pl,ind,ir,hi

$$g_i(\bar{y}_i(\boldsymbol{x}), y_i) = \mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i, y_i) \tag{22.4.13}$$

for some functions $\{\mathsf{h}_i(\ell, y_i)\}$. Table 22.2 lists some examples. In these problems the negative log-likelihood (22.4.6) has the following gradients:

$$\begin{aligned} \nabla^{[2,0]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{A}' \,\mathsf{diag}\left\{ \mathsf{h}_i{}^{[2,0]}([\boldsymbol{A}\boldsymbol{x}]_i, y_i) \right\} \boldsymbol{A} \\ \nabla^{[1,1]} \mathsf{L}(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{A}' \,\mathsf{diag}\left\{ \mathsf{h}_i{}^{[1,1]}([\boldsymbol{A}\boldsymbol{x}]_i, y_i) \right\}. \end{aligned}$$

Combining with (22.4.2) leads to the following expression for the gradient of penalized-likelihood image reconstruction methods:

$$\begin{aligned} \boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) &= \mathsf{diag}\left\{ -\mathsf{h}_i{}^{[1,1]}([\boldsymbol{A}\boldsymbol{x}]_i, y_i) \right\} \\ \boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y}) &= \mathsf{diag}\left\{ \mathsf{h}_i{}^{[2,0]}([\boldsymbol{A}\boldsymbol{x}]_i, y_i) \right\} \qquad (22.4.14) \\ \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) &\approx \left[\boldsymbol{A}' \boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{A} + \mathsf{R}(\boldsymbol{x}) \right]^{-1} \boldsymbol{A}' \boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) \Big|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})}. \qquad (22.4.15) \end{aligned}$$

For affine models $\bar{\boldsymbol{y}}(\cdot)$, this expression is exact except for any disregarded constraints such as nonnegativity.

s,srp,pl,ind,ex

22.4.3 Examples

The following examples for typical image reconstruction problems reveal a common form for the local impulse response.

22.4.3.1 Penalized weighted least squares (PWLS)

Consider the linear gaussian model with $\bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ and $\mathsf{h}_i(\ell, y_i) = \frac{w_i}{2}(\ell - y_i)^2$, *i.e.*, a *penalized weighted least squares* (*PWLS*) estimator. Note that $g_i(z, y) = \mathsf{h}_i(z, y)$ and $g_i{}^{[1,0]}(z, y_i) = w_i(z - y_i) = 0$ when $z = y_i$, as in (22.4.8) Combining (22.2.1) with (22.4.15) yields the following expression for the local impulse response:

e,srp,pl,ind,lir,pwls

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) = \left[\mathsf{F} + \mathsf{R}(\hat{\boldsymbol{x}}(\boldsymbol{y})) \right]^{-1} \mathsf{F} \boldsymbol{e}_j, \tag{22.4.16}$$

where $\boldsymbol{W} = \mathsf{diag}\{w_i\}$ and $\mathsf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is the *Fisher information matrix* associated with the model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{W}^{-1}\right)$. This is an exact expression because $\frac{\partial^2}{\partial x_j \partial x_k} \bar{y}_i(\boldsymbol{x}) = 0$ as explained below (22.4.3).

22.4.3.2 Quadratically penalized weighted least squares (QPWLS)

As a preview to why "choosing the regularization parameter" is considered challenging by some, consider the simplest case of white noise where $w_i = 1/\sigma^2$ and a quadratic penalty $\mathsf{R}(\boldsymbol{x}) = \beta \frac{1}{2} \boldsymbol{x}' \mathsf{R} \boldsymbol{x}$. Then (22.4.16) simplifies to

$$\bar{\boldsymbol{l}}^{(j)} = \left[\boldsymbol{A}'\boldsymbol{A} + \sigma^2 \beta \mathsf{R} \right]^{-1} \boldsymbol{A}' \boldsymbol{A} \boldsymbol{e}_j.$$

So the regularization parameter β effectively is *scaled* by the noise variance σ^2. This type of noise-dependent resolution effect is exacerbated further in more complicated noise models. §22.10 describes regularization design methods that address this issue.

22.4.3.3 Poisson emission penalized likelihood

For the emission tomography problem with Poisson measurement noise where $\bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}$ and $\mathsf{h}_i(\ell, y_i) = \ell + r_i - y_i \log(\ell + r_i)$, combining (22.2.1) with (22.4.15) yields the following expression for the local impulse response:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) = \left[\boldsymbol{A}' \,\mathsf{diag}\left\{ \frac{y_i}{\bar{y}_i^2(\boldsymbol{x})} \right\} \boldsymbol{A} + \mathsf{R}(\boldsymbol{x}) \right]^{-1} \boldsymbol{A}' \,\mathsf{diag}\left\{ \frac{1}{\bar{y}_i(\boldsymbol{x})} \right\} \boldsymbol{A}\boldsymbol{e}_j \Bigg|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})} . \tag{22.4.17}$$

e,srp,pl,lir,epl

This expression is approximate because one usually enforces nonnegativity in emission tomography. Typically the fitted model agrees with the data reasonably well, *i.e.*, $\bar{y}_i(\hat{\boldsymbol{x}}(\boldsymbol{y})) \approx y_i$, so the following "*plug in*" approximation is also useful [14]:

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) \approx \left[\boldsymbol{A}'\boldsymbol{D}(\boldsymbol{y})\boldsymbol{A} + \mathsf{R}(\boldsymbol{x}) \right]^{-1} \boldsymbol{A}'\boldsymbol{D}(\boldsymbol{y})\boldsymbol{A}\boldsymbol{e}_j, \tag{22.4.18}$$

e,srp,pl,lir,epl,approx

where $\boldsymbol{D}(\boldsymbol{y}) = \mathsf{diag}\left\{ \frac{1}{\max(y_i, 1)} \right\}$. Other approximations for $\boldsymbol{D}$ have also been investigated [15, 16].

22.4.3.4 Poisson transmission penalized likelihood

For the monoenergetic transmission tomography problem $\bar{y}_i(\boldsymbol{x}) = b_i \,\mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}]_i} + r_i$ and for Poisson measurement noise $\mathsf{h}_i(\ell, y_i) = b_i \,\mathrm{e}^{-\ell} + r_i - y_i \log\left(b_i \,\mathrm{e}^{-\ell} + r_i\right)$; combining (22.2.1) and (22.4.15) yields the following expression for the local impulse response:

$$\begin{aligned}
\boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) &= -\,\mathsf{diag}\left\{ 1 - \frac{r_i}{\bar{y}_i(\boldsymbol{x})} \right\} \\
\boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y}) &= \mathsf{diag}\left\{ \left[1 - \frac{y_i r_i}{\bar{y}_i^2(\boldsymbol{x})} \right] (r_i - \bar{y}_i(\boldsymbol{x})) \right\} \\
\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) &= \left[\boldsymbol{A}'\boldsymbol{D}_2(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{A} + \mathsf{R}(\boldsymbol{x}) \right]^{-1} \boldsymbol{A}'\boldsymbol{D}_1(\boldsymbol{x}, \boldsymbol{y}) \,\mathsf{diag}\left\{ -b_i \,\mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}]_i} \right\} \boldsymbol{A}\boldsymbol{e}_j \Big|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})} .
\end{aligned} \tag{22.4.19}$$

Invoking the assumption $\bar{y}_i(\hat{\boldsymbol{x}}(\boldsymbol{y})) \approx y_i$, this simplifies to an expression of the form (22.4.18) with

$$\boldsymbol{D}(\boldsymbol{y}) = \mathsf{diag}\left\{ \frac{(y_i - r_i)^2}{\max(y_i, 1)} \right\} .$$

22.5 Computing the local impulse response (s,srp,comp)

s,srp,comp

The preceding section described analytical approximations for the local impulse response of estimators defined implicitly as the minimizer of some cost function. This section describes methods for evaluating those approximations numerically.

22.5.1 Perturbation approximation

To evaluate the local impulse response of an estimator $\hat{\boldsymbol{x}}(\boldsymbol{y})$ for some measurement vector $\boldsymbol{y}$ (or for $\bar{\boldsymbol{y}}$), one can simply choose a small $\varepsilon > 0$ and compute

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}; \boldsymbol{x}) \approx \frac{\hat{\boldsymbol{x}}(\boldsymbol{y} + \varepsilon \boldsymbol{p}(\boldsymbol{x})) - \hat{\boldsymbol{x}}(\boldsymbol{y})}{\varepsilon}, \tag{22.5.1}$$

e,srp,comp,ljyx

where $\boldsymbol{p}(\boldsymbol{x}) \triangleq \nabla \bar{\boldsymbol{y}}(\boldsymbol{x})\, \boldsymbol{e}_j$ for computing (22.2.1), or $\boldsymbol{p}(\boldsymbol{x}) \triangleq \Delta\, \bar{\boldsymbol{y}}(f, \vec{\mathrm{x}})$ for computing (22.2.6). In other words, we reconstruct two images: from the original data $\boldsymbol{y}$, and one from modified data with an additional contribution due to a perturbing object point. Either of these cases can be applied even to real data where $\boldsymbol{x}_{\mathrm{true}}$ is unknown. In either case, the approximation accuracy will depend on the quality of the system model.

22.5.2 Monte Carlo estimator

For computing the local impulse response defined in (22.2.4) when no analytical expression for the estimator mean $\boldsymbol{\mu}(\boldsymbol{x})$ is known, one must resort to Monte Carlo methods. In the context of computer simulations, one can estimate $\boldsymbol{\mu}(\boldsymbol{x})$ by performing multiple realizations of the measurements $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M$, reconstructing $\hat{\boldsymbol{x}}$ from each, and computing the sample mean to yield $\hat{\boldsymbol{\mu}}(\boldsymbol{x}) = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{x}}(\boldsymbol{y}^m)$. Then one can repeat this process for measurements that are perturbed by the effect of a small point object, to yield $\hat{\boldsymbol{\mu}}(\boldsymbol{x} + \varepsilon \boldsymbol{e}_j)$. Then one can estimate the local impulse response using

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{x}) \approx \frac{\hat{\boldsymbol{\mu}}(\boldsymbol{x} + \varepsilon \boldsymbol{e}_j) - \hat{\boldsymbol{\mu}}(\boldsymbol{x})}{\varepsilon}.$$

This is an computationally demanding approach.

22.5.3 Unbiased estimator

An unbiased estimate for the local impulse response (22.2.4) was proposed in [14] by an insightful reviewer:

$$\bar{\boldsymbol{l}}_{\mathrm{unb}}^{(j)}(\boldsymbol{x}) \triangleq \frac{1}{M-1} \sum_{m=1}^{M} [\hat{\boldsymbol{x}}(\boldsymbol{y}^m) - \hat{\boldsymbol{\mu}}(\boldsymbol{x})] \frac{\partial}{\partial x_j} \log \mathsf{p}(\boldsymbol{y}^m; \boldsymbol{x}) \,.$$

See [17] for proof that this is unbiased. However, this is again relatively computationally demanding.

22.5.4 Matrix-based approximations

§22.4 showed that in many cases the local impulse response has the form (22.4.16). To compute this local impulse response, it would be inefficient to try to invert literally the $n_{\mathrm{p}} \times n_{\mathrm{p}}$ Hessian matrix. Instead, the practical approach is to apply an iterative algorithm such as *preconditioned conjugate gradients* (*PCG*) (see §14.6.2) to solve the following system of equations approximately:

e,srp,comp,system

$$[\mathbf{F} + \mathbf{R}] \, \bar{\boldsymbol{l}}^{(j)} = \mathbf{F} \boldsymbol{e}_j. \tag{22.5.2}$$

Computing this by PCG for a single pixel location j requires approximately the same work as iteratively reconstructing $\hat{\boldsymbol{x}}$. So usually it is feasible to perform for at most a small number of pixel locations. The next sections explore further approximations that can reduce computation.

IRT *The* `qpwls_pcg1.m` *routine is suitable when* $\mathbf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$.

22.6 Local shift-invariance (s,srp,local)

s,srp,local

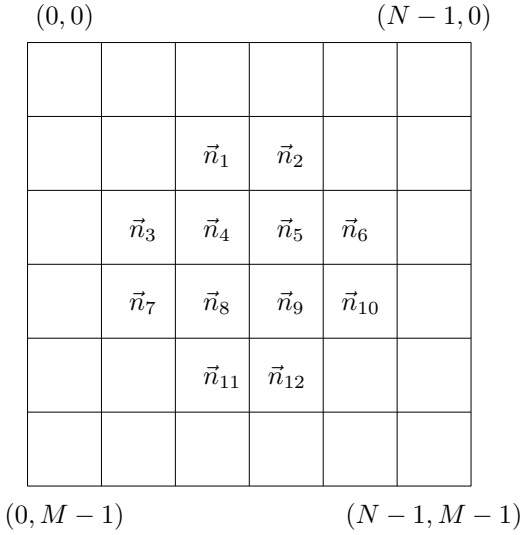
Local impulse response expressions such as (22.4.16) involve one or more $n_{\mathrm{p}} \times n_{\mathrm{p}}$ matrices. In general, these matrices are not Toeplitz and hence not circulant. However, often these problems exhibit approximate *local shift invariance*, as described in §4.4. We can exploit local shift invariance to accelerate computation of local impulse response functions and local noise properties.

Let $\boldsymbol{M}$ denote one of these $n_{\mathrm{p}} \times n_{\mathrm{p}}$ matrices, such as $\mathbf{F}$ or $\mathbf{R}$, or inverses of sums thereof. Then the matrix-vector operation $\boldsymbol{z} = \boldsymbol{M}\boldsymbol{x}$ can be expressed equivalently as

e,srp,local,super

$$z[\vec{n}] = 1_{\{\vec{n} \in \mathcal{S}\}} \sum_{\vec{m} \in \mathcal{S}} h(\vec{n}, \vec{m}) \, x[\vec{m}] = 1_{\{\vec{n} \in \mathcal{S}\}} \sum_{\vec{m}} h(\vec{n}, \vec{m}) \, 1_{\{\vec{m} \in \mathcal{S}\}} \, x[\vec{m}], \tag{22.6.1}$$

where $\mathcal{S} \triangleq \{\vec{n}_j : j = 1, \ldots, n_{\mathrm{p}}\}$ and $\vec{n}_j$ denotes the spatial location corresponding to the jth column of $\boldsymbol{M}$. In other words, the elements of $\boldsymbol{M}$ correspond to $M_{kj} = h(\vec{n}_k, \vec{n}_j)$. Fig. 22.6.1 illustrates an example of a set $\mathcal{S}$.



fig,n,m,square,lattice

Figure 22.6.1: Illustration of $N \times M$ image lattice with $n_{\mathrm{p}} = 12$ pixels within support mask to be estimated.

Near a given location $\vec{n}_0$ of interest, we define a *local impulse response* as follows:

e,src,loca,h0

$$h_0(\vec{n}) \triangleq h(\vec{n}_0 + \alpha\vec{n}, \vec{n}_0 - (1-\alpha)\,\vec{n}) 1_{\{\vec{n}_0 + \alpha\vec{n} \in \mathcal{S}\}} 1_{\{\vec{n}_0 - (1-\alpha)\vec{n} \in \mathcal{S}\}}, \tag{22.6.2}$$

where usually we take $\alpha = 1$. However, sometimes we can approximate h even for noninteger arguments, in which case $\alpha = 1/2$ may also be useful. The subscript "0" in $h_0(\vec{n})$ concisely reminds us of the dependence on location $\vec{n}_0$. There may be other definitions of $h_0(\vec{n})$ that lead to better approximations below. One possibility is to consider

optimal circulant approximations to Toeplitz (and non-Toeplitz) matrices, *e.g.*, [18–20]. Investigating such alternatives is an *open problem.*

If the operator $\boldsymbol{M}$ is approximately shift invariant, then we can approximate the superposition sum (22.6.1) by (almost) a convolution sum:

e,srp,local,approx

$$z[\vec{n}] \approx 1_{\{\vec{n} \in \mathcal{S}\}} \sum_{\vec{m}} h_0(\vec{n} - \vec{m}) \, 1_{\{\vec{m} \in \mathcal{S}\}} \, x[\vec{m}] \,. \tag{22.6.3}$$

This approximation should be accurate provided $\vec{n}$ and $\vec{m}$ are "sufficiently close" to $\vec{n}_0$ relative to the width of h_0. This expression is *almost* a convolution sum, except for the "edge conditions" of the indicator functions. If the point $\vec{n}_0$ is not "too close" to the boundaries of the support mask $\mathcal{S}$, then we can disregard the indicator functions and treat the expression as a convolution.

Define the matrix $\boldsymbol{M}_0$ by $[\boldsymbol{M}_0]_{kj} = h_0(\vec{n}_k - \vec{n}_j)$. Then this matrix represents the linear operation in (22.6.3). However, in general the structure of this matrix is quite complicated due to the set $\mathcal{S}$ illustrated in Fig. 22.6.1. To characterize $\boldsymbol{M}_0$, it is helpful to first let $\boldsymbol{T}$ be the $NM \times n_\mathrm{p}$ matrix such that

$$T_{1+n+mN,j} = \begin{cases} 1, & \vec{n}_j = (n,m) \\ 0, & \text{otherwise,} \end{cases}$$

for $n = 0, \dots, N-1$, $m = 0, \dots, M-1$ and $j = 1, \dots, n_\mathrm{p}$. The purpose of $\boldsymbol{T}$ is to *embed* the n_p elements of $\boldsymbol{x}$, as shown in Fig. 22.6.1, back to the 2D $N \times M$ lattice. Then $\boldsymbol{M}_0 = \boldsymbol{T}' \hat{\boldsymbol{M}}_0 \boldsymbol{T}$, where $\left[\hat{\boldsymbol{M}}_0\right]_{\vec{n},\vec{n}'} = h_0(\vec{n} - \vec{n}')$ is an $NM \times NM$ matrix that is *block Toeplitz with Toeplitz blocks* (*BTTB*), at least if we disregard the indicator functions in (22.6.2). Thus we can make a circulant approximation to $\hat{\boldsymbol{M}}_0$ [21], as described in the next section. Such approximations are often reasonably accurate except near the edges of the FOV, where the differences between "Toeplitz"and "circulant" end conditions are largest. We have described the matrix $\boldsymbol{T}$ for a 2D example, but more generally we need a $N_\mathrm{p} \times n_\mathrm{p}$ matrix where N_p is the number of pixels in the rectangular lattice that is needed for taking the DFT of the appropriate dimension. In 2D we have $N_\mathrm{p} = NM$.

IRT *The routine* `embed.m` *performs the role of* $\boldsymbol{T}$.

The next section uses such circulant approximations to accelerate computation of the *local impulse response* (22.4.16).

22.7 Local Fourier approximations (s,srp,dft)

s,srp,dft

As described in §4.4, many image reconstruction methods are approximately locally shift invariant. For such methods, we can apply local Fourier analysis to simplify greatly the computation of local impulse response functions. We focus here on local impulse response functions of the form

e,srp,dft,lj,mat

$$\bar{\boldsymbol{l}}^{(j)} = [\mathsf{F} + \mathsf{R}]^{-1} \mathsf{F} \boldsymbol{e}_j, \tag{22.7.1}$$

where typically $\mathsf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ and $\boldsymbol{W}$ is some weighting matrix, possibly dependent on $\boldsymbol{y}$, and R is the roughness penalty Hessian, possibly dependent on $\boldsymbol{x}$. §22.4 showed that many problems are of this form.

The key idea is that $\mathsf{F}\boldsymbol{e}_j$ and $\mathsf{R}\boldsymbol{e}_j$ are usually "concentrated" around the jth pixel. And furthermore we often find that the operator F is locally shift invariant, meaning that $\mathsf{F}\boldsymbol{e}_j$ and $\mathsf{F}\boldsymbol{e}_{j'}$ are similar except for the shift from the jth to the j'th pixel. Likewise for the operator R.

Shift invariant operators can be represented exactly by Toeplitz matrices, and *locally* shift invariant operators can be approximated by circulant matrices. This idea was perhaps first used in the context of tomography by Clinthorne *et al.* [22] for developing preconditioners (see also [23]), and was used later for spatial resolution analyses [14,15,24,25].

Consider the 2D case of $N \times M$ images. Let $\mathsf{F}_k^{(j)}$ denote the $N \times M$-point DFT coefficients for the vector $\boldsymbol{T}\mathsf{F}\boldsymbol{e}_j$, and let $\mathsf{R}_k^{(j)}$ denote the DFT coefficients for the vector $\boldsymbol{T}\mathsf{R}\boldsymbol{e}_j$. Then for $j' \approx j$ the locally circulant approximation is:

e,srp,dft,F

$$\boldsymbol{T}\mathsf{F}\boldsymbol{e}_{j'} \approx \boldsymbol{Q}^{-1} \operatorname{diag}\left\{\mathsf{F}_k^{(j)}\right\} \boldsymbol{Q}\boldsymbol{T}\boldsymbol{e}_{j'}, \tag{22.7.2}$$

where $\boldsymbol{Q}$ denotes the $N \times M$-point DFT matrix for 2D images defined in (1.4.31). (See [15] for the 3D case.) Similarly,

$$\boldsymbol{T}\mathsf{R}\boldsymbol{e}_{j'} \approx \boldsymbol{Q}^{-1} \operatorname{diag}\left\{\mathsf{R}_k^{(j)}\right\} \boldsymbol{Q}\boldsymbol{T}\boldsymbol{e}_{j'}.$$

We combine these approximations as follows:

$$\boldsymbol{T}\,[\mathsf{F} + \mathsf{R}]^{-1} \approx \boldsymbol{Q}^{-1} \operatorname{diag}\left\{\mathsf{F}_k^{(j)} + \mathsf{R}_k^{(j)}\right\}^{-1} \boldsymbol{Q}\boldsymbol{T}.$$

Combining these approximations with (22.7.1) yields the following *local Fourier approximation* for the local impulse response (22.7.1):

e,srp,dft,lj,dft

$$\boldsymbol{T}\,\bar{\boldsymbol{l}}^{(j)} \approx \boldsymbol{Q}^{-1} \operatorname{diag}\left\{\frac{\mathsf{F}_k^{(j)}}{\mathsf{F}_k^{(j)} + \mathsf{R}_k^{(j)}}\right\} \boldsymbol{Q}\boldsymbol{T}\boldsymbol{e}_j. \tag{22.7.3}$$

One can compute this approximation with just a few FFT operations, so it is quite practical.

There is one subtle point in computing the $\mathsf{F}_k^{(j)}$ and $\mathsf{R}_k^{(j)}$ values. Assuming that $\boldsymbol{W}$ is symmetric positive semidefinite (*e.g.*, diagonal with nonnegative elements), the matrix $\mathbf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is Hermitian positive semidefinite, so it is only natural to require its circulant approximation to also be Hermitian positive semidefinite. This means that the $\mathsf{F}_k^{(j)}$ values should be nonnegative real numbers when $\boldsymbol{e}_j$ corresponds to the pixel at the center of the field of view. In practice, if the FFT operation yields complex or negative values, we take the real part of $\mathsf{F}_k^{(j)}$ and set the negative elements to zero. Similarly for the R_k values in the usual case where $\mathbf{R}$ is positive semidefinite. When $\mathsf{F}_k^{(j)}$ and $\mathsf{R}_k^{(j)}$ are so treated, the denominator of (22.7.3) will be real positive numbers.

IRT *See* `qpwls_psf.m` *for examples.*

22.8 Complex data, real images (s,srp,complex)

s,srp,complex

When the data $\boldsymbol{y}$ and the system model $\boldsymbol{A}$ are complex valued, but the image $\boldsymbol{x}$ is real valued, such as in certain MRI applications, then the gradient expression (22.4.15) must be modified. For simplicity, consider the QPWLS problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{p}}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}^{1/2}} + \beta \|\boldsymbol{C}\boldsymbol{x}\| = [\mathrm{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\} + \beta \boldsymbol{C}'\boldsymbol{C}]^{-1} \mathrm{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{y}\}$$

for which

$$\mathsf{E}[\hat{\boldsymbol{x}}] = [\mathrm{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\} + \beta \boldsymbol{C}'\boldsymbol{C}]^{-1} \mathrm{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\}\, \boldsymbol{x}.$$

In this case, the general local impulse form (22.7.1) still holds for $\mathbf{F} = \mathrm{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\}$, but the local frequency response values F_k in (22.7.2) must be modified due to the $\mathrm{real}\{\}$ operator. Roughly, the usual F_k must be replaced by $\frac{1}{2}(\mathsf{F}_k + \mathsf{F}_{-k})$ where " F_{-k}" denotes the local frequency response value at the mirror image location in frequency space. In 1D we have $\mathsf{F}_{-k} = \mathsf{F}_{N-k}$. See also Example 28.3.2 and [26].

22.9 Fisher information approximations (s,srp,fish)

s,srp,fish

In several of the cases described in §22.4.3, the matrix $\mathbf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is the *Fisher information*, or an approximation thereof, for estimating $\boldsymbol{x}$ from $\boldsymbol{y}$. This matrix has a central role in the analysis of both spatial resolution and noise properties of image reconstruction methods based on estimation principles. Although one can use FFTs as described in §22.7 to compute local impulse response functions for a few spatial locations of interest, if one wants to compute them for many spatial locations (*e.g.*, for regularization design) then even the DFT approach can become impractical. This has motivated the development of several approximations to $\mathbf{F}$ that facilitate rapid computation of the local impulse response.

22.9.1 Certainty approximation

In some applications such as emission tomography, the system matrix $\boldsymbol{A}$ can be factored as follows:

$$\boldsymbol{A} = \mathsf{diag}\{c_i\}\, \boldsymbol{G}\, \mathsf{diag}\{s_j\},$$

where the c_i values are ray-dependent factors such as detector efficiencies, PET attenuation factors, etc., the s_j values are pixel-dependent factors such as spatial sensitivity variations and (in SPECT) "first order" attenuation factors (*cf.* the image-space Chang method [27] for SPECT attenuation correction), and $\boldsymbol{G}$ denotes an object-independent, shift-invariant system model. When this factorization is applicable, the Fisher information "simplifies" as follows:

$$\mathbf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} = \mathsf{diag}\{s_j\}\, \boldsymbol{G}'\, \mathsf{diag}\{c_i^2 w_i\}\, \boldsymbol{G}\, \mathsf{diag}\{s_j\},$$

assuming $\boldsymbol{W} = \mathsf{diag}\{w_i\}$. Now consider the approximation

$$\boldsymbol{G}'\, \mathsf{diag}\{c_i^2 w_i\}\, \boldsymbol{G} \approx \mathsf{diag}\{\kappa_j\}\, \boldsymbol{G}'\boldsymbol{G}\, \mathsf{diag}\{\kappa_j\}, \tag{22.9.1}$$

e,srp,fish,approx1

where we define the following "certainty" factors

$$\kappa_j = \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} g_{ij}^2 c_i^2 w_i}{\sum_{i=1}^{n_\mathrm{d}} g_{ij}^2}}. \tag{22.9.2}$$

e,srp,fish,kapj

This leads to the following certainty-based Fisher information approximation:

$$\mathbf{F} \approx \mathsf{diag}\{s_j \kappa_j\}\, \boldsymbol{G}'\boldsymbol{G}\, \mathsf{diag}\{s_j \kappa_j\}\,. \tag{22.9.3}$$

e,srp,fish,cert

We chose the κ_j factors in (22.9.2) so that this approximation is *exact* along its diagonal, which is usually where the largest elements of $\mathbf{F}$ are located due to the $1/r$ response of tomographic systems.

Substituting (22.9.3) into (22.4.16) and simplifying yields the following local impulse response approximation:

$$\begin{aligned}
\bar{\boldsymbol{l}}^{(j)} &\approx \left[\mathsf{diag}\{s_j\kappa_j\}\,\boldsymbol{G}'\boldsymbol{G}\,\mathsf{diag}\{s_j\kappa_j\} + \mathsf{R}\right]^{-1} \left(\mathsf{diag}\{s_j\kappa_j\}\,\boldsymbol{G}'\boldsymbol{G}\,\mathsf{diag}\{s_j\kappa_j\}\right)\boldsymbol{e}_j \\
&= \mathsf{diag}\left\{\frac{1}{s_j\kappa_j}\right\}\left[\boldsymbol{G}'\boldsymbol{G} + \mathsf{diag}\left\{\frac{1}{s_j\kappa_j}\right\}\mathsf{R}\,\mathsf{diag}\left\{\frac{1}{s_j\kappa_j}\right\}\right]^{-1}\boldsymbol{G}'\boldsymbol{G}\,(s_j\kappa_j)\,\boldsymbol{e}_j \qquad (22.9.4) \\
&\approx \left[\boldsymbol{G}'\boldsymbol{G} + \frac{1}{s_j^2\kappa_j^2}\mathsf{R}\right]^{-1}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_j. \qquad (22.9.5)
\end{aligned}$$

The final approximation is reasonable when the regularization strength is weak enough that the local impulse response is narrow. In other words, the certainty approximation suggests that the local impulse response for the jth pixel corresponds to a penalized LS estimator with system matrix $\boldsymbol{G}$ and with a regularization parameter that is scaled down by $s_j^2\kappa_j^2$. This approximation captures how the width of the local impulse response changes with the noise statistics, but does not accurately model the *shape* of the local impulse response [28].

The g_{ij}^2 terms in (22.9.2) are inconvenient for many projection / back-projections methods, so in practice we usually use the following approximation instead:

$$\kappa_j \approx \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} g_{ij} c_i^2 w_i}{\sum_{i=1}^{n_\mathrm{d}} g_{ij}}}. \qquad (22.9.6)$$

e,srp,fish,kapj,1

It is very simple to compute these κ_j values prior to iterating, as follows.

- First determine the factors $\tilde{w}_i = w_i c_i^2$ in sinogram space using the statistics of the measurements.
- Compute $\sum_{i=1}^{n_\mathrm{d}} g_{ij}\tilde{w}_i$ by backprojecting the $\tilde{w}_i$ "sinogram data" into image space.
- Compute $\sum_{i=1}^{n_\mathrm{d}} g_{ij}$ by backprojecting a "sinogram data" array full of the value 1.0 into image space. This factor does not depend on the patient, so it can be precomputed and stored for a given system geometry. (It does depend on the selected FOV.)
  In parallel beam CT, usually $\sum_{i=1}^{n_\mathrm{d}} g_{ij}$ is a scalar constant. In fan beam CT it probably depends on the pixel's distance from isocenter in 2D and it may depend on other factors such as the pitch in helical CT scans.
- Divide those two image space arrays over the FOV of interest, and take the square root.

For efficient implementation, it is quite reasonable to use a highly angularly downsampled backprojector for computing the κ_j factors. So computer κ_j can add very minimal computation time prior to iterating.

Here is a block diagram of the data flow.

Sinogram data → w_i → Angularly downsampled backprojection → κ_j → Regularizer $\mathsf{R}(\boldsymbol{x})$

Fig. 22.9.1 compares the "exact" formula (22.9.2) with the approximation (22.9.6) for the case of CT imaging of the image shown [29]. For the approximation (22.9.6), I downsampled by a factor of 10 along the angular dimension. The profile illustrates that (22.9.6) is a very good approximation even with substantial down sampling. See Problem 22.4.

take limit as $\mathrm{d}r \to 0$ to prove that not squaring is reasonable!

illustrate (22.9.1) using `fig_srp_kappa_approx1.m`

22.9.2 Angular dependent approximation

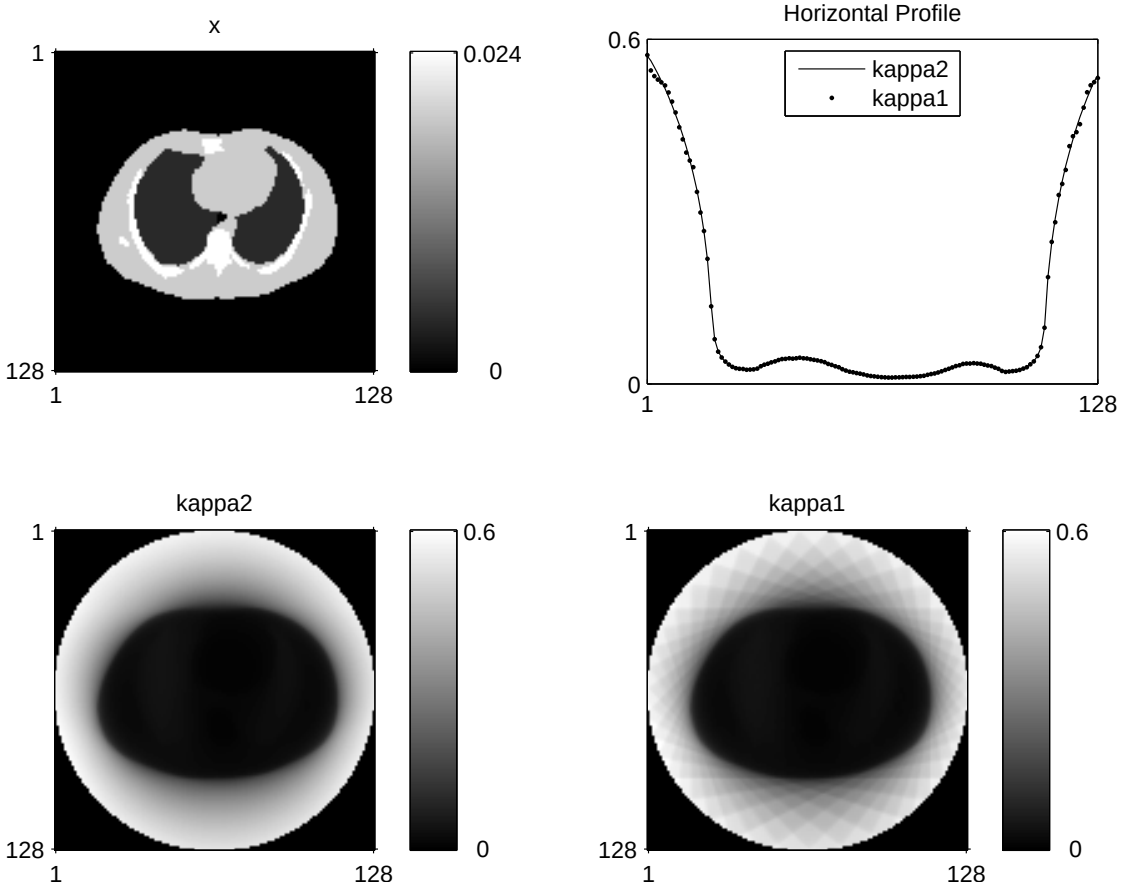
In the preceding analysis, variations in certainty as a function of projection angle are not captured because the index "i" sweeps over both radial and angular position. To form a more accurate approximation, here we replace i with lm, where $l = 1, \ldots, L$ is the radial position index, and $m = 1, \ldots, M$ is the projection angle index. (This is for 2D image reconstruction; one can generalize easily to the 3D case.) So a_{lmj} denotes the system matrix element for the contribution of the jth pixel to the lth radial position and the mth angular position in the sinogram. With this notation, the Fisher information matrix can be approximated as follows [30]:

$$\begin{aligned}
F_{jk} &= s_j \left[\boldsymbol{G}'\,\mathsf{diag}\left\{c_i^2 w_i\right\}\boldsymbol{G}\right]_{jk} s_k \\
&= s_j s_k \sum_m \sum_l g_{lmj} c_{i(l,m)}^2 w_{i(l,m)} g_{lmk} \\
&\approx s_j s_k \sum_m w_{mj} \sum_l g_{lmj} g_{lmk}
\end{aligned}$$

where $i(l,m) = (m-1)L + l$ denotes the measurement index corresponding to indices l, m, and where we define the following angle-dependent certainty factors:

$$w_{mj} \triangleq \frac{\sum_l g_{lmj}^2 c_{i(l,m)}^2 w_{i(l,m)}}{\sum_l g_{lmj}^2}. \qquad (22.9.7)$$

e,srp,fish,wmj



fig_srp_fish_kappa1

Figure 22.9.1: Empirical comparison of (22.9.2) "kappa2" with the approximation (22.9.6) "kappa1."

Again, this approximation is exact along the diagonals of the Fisher information matrix $\mathbf{F}$. It is also exact if $c_{i(l,m)}$ and $w_{i(l,m)}$ depend only on the projection angle m, rather than on the radial position l. This is rarely exactly the case, but is often a very close approximation. In matrix form:

$$\mathbf{F}\boldsymbol{e}_j \approx s_j^2 \sum_{m=1}^{M} w_{mj} \boldsymbol{G}_m' \boldsymbol{G}_m \boldsymbol{e}_j,$$

where $\boldsymbol{G}_m$ denotes the $L \times n_\mathrm{p}$ matrix with elements g_{lmj} for $l = 1, \dots, L$ and $j = 1, \dots, n_\mathrm{p}$.

The angular certainty factors (22.9.7) are useful for regularization design; see §5.1.

For further approximations and their use in the design of regularization methods, see [24, 25, 31].

s,srp,design

22.10 Regularization design based on spatial resolution properties (s,srp,design)

Using the type of spatial resolution analysis described in this chapter, several methods have been proposed for designing modified regularization methods. Some methods have focused on achieving uniform spatial resolution *e.g.*, [14, 24, 25, 30], others on uniform contrast recovery, *e.g.*, [15]. In general the design of *quadratic* regularization methods has been the most successful; the nonquadratic case is largely an *open problem*, although there has been recent progress using *local perturbation* analysis [32, 33].

22.10.1 Certainty-based design

The earliest and simplest approach [14,34] used the certainty-based Fisher information approximation (22.9.3) and the corresponding local impulse response approximation (22.9.4) to propose the following modified quadratic roughness penalty function:

$$\mathsf{R}(\boldsymbol{x}) = \beta \sum_{j=1}^{n_\mathrm{p}} \sum_{k \in \mathcal{N}_j} \kappa_j \kappa_k \, \psi(x_j - x_k),$$

where $\psi(t) = t^2/2$, the κ_j values were defined in (22.9.2), and $\mathcal{N}_j$ denotes a neighborhood of the jth pixel. This very simple modified penalty yields approximately uniform *average* spatial resolution at each pixel, but the local impulse response can still be very anisotropic [14].

IRT *See* `Reg1.m` *for an example.*

22.10.2 Analytical Fourier-based design

Regularization designs based on analytical approximations have also been successful; see [31, 35] and §5.1. Wang *et al.* designed regularizer coefficients to induce a certain desired frequency spectrum [36].

22.11 Resolution analysis of iterations (s,srp,iter)

s,srp,iter

The preceding analyses have considered estimators that have been iterated until convergence to a minimizer of a cost function. For iterative reconstruction methods that are not based on a cost function, it may still be of interest to examine the spatial resolution properties. The material in this section was inspired by the work of Mustafovic and Thielemans [37].

Consider an iteration of the form

e,srp,iter,g

$$\boldsymbol{x}^{(n+1)} = \mathcal{M}(\boldsymbol{x}^{(n)}, \boldsymbol{y}), \tag{22.11.1}$$

and assume that this iteration converges to a limit $\hat{\boldsymbol{x}}(\boldsymbol{y})$. To analyze the spatial resolution properties of $\hat{\boldsymbol{x}}$, once again the key step is to find $\nabla \hat{\boldsymbol{x}}(\boldsymbol{y})$.

If the iteration converges to a limit $\hat{\boldsymbol{x}}(\boldsymbol{y})$, then that limit must satisfy the following fixed-point relationship

e,srp,iter,fixed

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \mathcal{M}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}). \tag{22.11.2}$$

Taking the row gradient of both sides with respect to $\boldsymbol{y}$ yields

$$\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = \nabla^{[1,0]} \mathcal{M}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) + \nabla^{[0,1]} \mathcal{M}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}).$$

Rearranging and "solving" yields

$$\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) = \left[\boldsymbol{I} - \nabla^{[1,0]} \mathcal{M}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) \right]^{-1} \nabla^{[0,1]} \mathcal{M}(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}),$$

assuming that $\boldsymbol{I} - \nabla^{[1,0]} \mathcal{M}$ is invertible.

As an example, consider a "smoothed" diagonally preconditioned gradient descent iteration of the form

e,srp,g,pgd

$$\mathcal{M}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{S} \left(\boldsymbol{x} - \boldsymbol{D}(\boldsymbol{x}) \nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \right) \tag{22.11.3}$$

for some $n_\mathrm{p} \times n_\mathrm{p}$ smoothing matrix $\boldsymbol{S}$ (possibly the identity matrix) and some preconditioning matrix $\boldsymbol{D}$. Then

$$\nabla^{[0,1]} \mathcal{M}(\boldsymbol{x}, \boldsymbol{y}) = -\boldsymbol{S} \boldsymbol{D}(\boldsymbol{x}) \nabla^{[1,1]} \Psi(\boldsymbol{x}, \boldsymbol{y})$$

and

$$\nabla^{[1,0]} \mathcal{M}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{S} \left[\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{x}) \nabla^{[2,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) - \sum_{j=1}^{n_\mathrm{p}} \boldsymbol{e}_j \nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \, \nabla D_j'(\boldsymbol{x}) \right],$$

where D_j denotes the jth row of $\boldsymbol{D}$. Suppose $\boldsymbol{D}$ is diagonal with jth diagonal element $d_j(x_j)$. Then

$$\sum_{j=1}^{n_\mathrm{p}} \boldsymbol{e}_j \nabla^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \, \nabla D_j'(\boldsymbol{x}) = \mathsf{diag}\left\{ \dot{d}_j(x_j) \nabla_j^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \right\}.$$

Combining we have

$$\begin{aligned} \nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) \quad = \quad & \left[\boldsymbol{I} - \boldsymbol{S} \left(\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{x}) \nabla^{[2,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) - \mathsf{diag}\left\{ \dot{d}_j(x_j) \nabla_j^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \right\} \right) \right]^{-1} \\ & \boldsymbol{S} \boldsymbol{D}(\boldsymbol{x}) \left[-\nabla^{[1,1]} \Psi(\boldsymbol{x}, \boldsymbol{y}) \right] \Big|_{\boldsymbol{x} = \hat{\boldsymbol{x}}(\boldsymbol{y})}. \end{aligned}$$

Substituting into (22.2.6) yields an exact expression for the local impulse response. Ignoring the nonnegativity constraint, if the filtering is modest, then the gradient of the cost function should be zero or nearly zero at the limit $\hat{\boldsymbol{x}}(\boldsymbol{y})$. Thus we assume the $\nabla_j^{[1,0]} \Psi(\boldsymbol{x}, \boldsymbol{y})$ terms are zero for $\boldsymbol{x} \approx \hat{\boldsymbol{x}}(\boldsymbol{y})$. This leads to the simplified expression:

$$\nabla \hat{\boldsymbol{x}}(\boldsymbol{y}) \approx \left[\boldsymbol{I} - \boldsymbol{S} \left(\boldsymbol{I} - \boldsymbol{D}(\hat{\boldsymbol{x}}) \nabla^{[2,0]} \Psi(\hat{\boldsymbol{x}}, \boldsymbol{y}) \right) \right]^{-1} \boldsymbol{S} \boldsymbol{D}(\hat{\boldsymbol{x}}) \left[-\nabla^{[1,1]} \Psi(\hat{\boldsymbol{x}}, \boldsymbol{y}) \right].$$

Substituting into (22.2.6) yields an approximate expression for the local impulse response.

In particular, if $\boldsymbol{S} = \boldsymbol{I}$, then

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}) \approx \left[\nabla^{[2,0]}\,\Psi(\hat{\boldsymbol{x}}, \boldsymbol{y})\right]^{-1} \left[-\nabla^{[1,1]}\,\Psi(\hat{\boldsymbol{x}}, \boldsymbol{y})\right] \boldsymbol{p},$$

an expression that is familiar from §22.4 and [24, 28].

As a concrete example, consider the *EMS algorithm* [38], in which

$$\begin{aligned}
\Psi(\boldsymbol{x}, \boldsymbol{y}) &= (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{r})'\boldsymbol{1} - \boldsymbol{y}' \log(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}) \\
\nabla^{[1,0]}\,\Psi(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{A}'\boldsymbol{1} - \boldsymbol{A}'(\boldsymbol{y} \oslash \bar{\boldsymbol{y}}(\boldsymbol{x})) \\
\nabla^{[2,0]}\,\Psi(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{A}'\,\mathsf{diag}\left\{y_i / \bar{y}_i^2(\boldsymbol{x})\right\} \boldsymbol{A} \\
-\nabla^{[1,1]}\,\Psi(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{A}'\,\mathsf{diag}\{1 / \bar{y}_i(\boldsymbol{x})\} \\
\boldsymbol{D}(\boldsymbol{x}) &= \mathsf{diag}\{x_j / a_j\},
\end{aligned}$$

where $a_j = \sum_{i=1}^{n_\mathrm{d}} a_{ij} = [\boldsymbol{A}'\boldsymbol{1}]_j$. Thus we have

$$\nabla\,\hat{\boldsymbol{x}}(\boldsymbol{y}) \approx \left[\boldsymbol{I} - \boldsymbol{S}\left(\boldsymbol{I} - \boldsymbol{D}(\hat{\boldsymbol{x}})\boldsymbol{A}'\,\mathsf{diag}\left\{y_i / \bar{y}_i^2(\hat{\boldsymbol{x}})\right\} \boldsymbol{A}\right)\right]^{-1} \boldsymbol{S}\boldsymbol{D}(\hat{\boldsymbol{x}})\boldsymbol{A}'\,\mathsf{diag}\{1 / \bar{y}_i(\hat{\boldsymbol{x}})\}\,.$$

Usually $y_i \approx \bar{y}_i(\hat{\boldsymbol{x}}(\boldsymbol{y}))$, so

$$\nabla\,\hat{\boldsymbol{x}}(\boldsymbol{y}) \approx \left[\boldsymbol{I} - \boldsymbol{S}\left(\boldsymbol{I} - \boldsymbol{D}(\hat{\boldsymbol{x}})\boldsymbol{A}'\,\mathsf{diag}\{1/y_i\}\,\boldsymbol{A}\right)\right]^{-1} \boldsymbol{S}\boldsymbol{D}(\hat{\boldsymbol{x}})\boldsymbol{A}'\,\mathsf{diag}\{1/y_i\}\,.$$

In particular, if $\boldsymbol{S} = \boldsymbol{I}$ (no smoothing) and $\hat{\boldsymbol{x}} > \boldsymbol{0}$, then the local impulse response (22.2.6) approximation simplifies to

$$\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}) \approx \left[\boldsymbol{A}'\,\mathsf{diag}\{1/y_i\}\,\boldsymbol{A}\right]^{-1} \boldsymbol{A}'\,\mathsf{diag}\{1/y_i\}\,\boldsymbol{p}.$$

When $\boldsymbol{p} = \boldsymbol{A}\boldsymbol{e}_j$, this further simplifies to $\bar{\boldsymbol{l}}^{(j)}(\boldsymbol{y}) \approx \boldsymbol{e}_j$. This result is consistent with the conventional wisdom that the unregularized (and unfiltered) EM algorithm converges to an image with approximately perfect spatial resolution.

Unfortunately, this analysis is inapplicable to typical OS-type algorithms because the form of such algorithms differs in important ways from (22.11.3). In fact, even (22.11.1) is inapplicable to OS-EM because the diagonal preconditioners used in OS-EM change with each subset, so we would need to write something like $\mathcal{M}^{(n)}(\boldsymbol{x}^{(n)}, \boldsymbol{y})$. Furthermore, most OS algorithms do not converge, so the limiting argument (22.11.2) is inapplicable.

Relaxed OS algorithms with subset-invariant diagonal preconditioning [39], if modified for some reason to include inter-iteration filtering, may still be amenable to this type of analysis.

22.12 PSF of denoising with quadratic regularization (s,srp,denoise)

s,srp,denoise

This section analyzes the spatial resolution properties of quadratically penalized least squares deblurring and denoising problems.

22.12.1 1D analysis of impulse response (s,srp,dtft)

s,srp,dtft

For simplicity, consider the case of a 1D discrete-time signal of infinite extent: $\boldsymbol{x} = \{x(n)\}\,,\ n \in \mathbb{Z}$. This type of analysis is described in [40]. Consider the measurement model

$$y(n) = g(n) * x(n) + \epsilon(n),$$

where $g(n)$ represents linear shift-invariant blur. A natural quadratically penalized least squares cost function for estimating $\boldsymbol{x}$ from $\{y(n)\}$ for this problem is the following:

$$\Psi(\boldsymbol{x}) = \sum_n \frac{1}{2} \left|y(n) - (g * x)(n)\right|^2 + \beta \sum_n \frac{1}{2} \left|(c * x)(n)\right|^2.$$

For example, if $c(n) = \delta[n] - \delta[n-1]$, then the penalty term simplifies to $\sum_n \frac{1}{2} \left|x(n) - x(n-1)\right|^2$. The simplest way to find the minimizer of this cost function is to first apply Parseval's theorem and work in the frequency domain:

$$\begin{aligned}
4\pi\,\Psi(\boldsymbol{x}) &= \int_{-\pi}^{\pi} \left|Y(\omega) - G(\omega)X(\omega)\right|^2 + \beta \left|C(\omega)X(\omega)\right|^2 \mathrm{d}\omega \\
&\equiv \int_{-\pi}^{\pi} \left(|G(\omega)|^2 + |C(\omega)|^2\right) \left|X(\omega) - \frac{G^*(\omega)Y(\omega)}{|G(\omega)|^2 + \beta|C(\omega)|^2}\right|^2 \mathrm{d}\omega,
\end{aligned}$$

by *completing the square*. Thus we have the following expression for the DTFT of the minimizer of $\Psi(\boldsymbol{x})$:

$$X(\omega) = \frac{G^*(\omega)}{|G(\omega)|^2 + \beta|C(\omega)|^2} Y(\omega).$$

Thus, the frequency response of this (linear, shift invariant estimator) is

$$H(\omega) = \frac{G^*(\omega)}{|G(\omega)|^2 + \beta |C(\omega)|^2}.$$

As a concrete example, consider the "denoising" problem where $g(n) = \delta[n]$ and hence $G(\omega) = 1$, and $c(n) = \delta[n] - \delta[n-1]$ so $C(\omega) = 1 - z^{-1}$, where $z = \mathrm{e}^{\imath\omega}$. Thus, in terms of the Z transform:

$$H(z) = \frac{1}{1 + \beta \left(1 - z^{-1}\right)\left(1 - z\right)} = \frac{1}{1 + 2\beta - \beta z^{-1} - \beta z}.$$

Applying a partial fraction expansion, we have

$$H(z) = \frac{1}{\sqrt{1 + 4\beta}} \left(\frac{1}{1 - r z^{-1}} - \frac{1}{1 - r^{-1} z^{-1}} \right),$$

where

$$r = \alpha - \sqrt{\alpha^2 - 1}, \qquad \alpha = 1 + \frac{1}{2\beta}.$$

Taking the inverse Z transform, we find that the impulse response is

$$h(n) = \frac{1}{\sqrt{1 + 4\beta}} \left(1 + \frac{1}{2\beta} - \sqrt{\frac{1}{\beta} + \frac{1}{4\beta^2}} \right)^{|n|}.$$

The parenthesized argument lies in the interval $(0, 1)$, so this impulse response is a double-sided decaying geometric series. Interestingly, this expression seems simpler than the solution to Problem 1.17.

22.12.2 N-D analysis

Now consider the N-D shift-invariant deblurring problem:

$$y(\vec{n}) = g(\vec{n}) * x(\vec{n}) + \varepsilon(\vec{n}),$$

where now "$\vec{n}$" denotes a point in $\mathbb{Z}^{\bar{d}}$ where $\bar{d}$ is the spatial dimensionality. In multi-dimensional problems we usually penalize roughness in several directions, so the a natural quadratically regularized least-squares cost function is

$$\Psi(\boldsymbol{x}) = \sum_{\vec{n}} \frac{1}{2} \left| y(\vec{n}) - (g * x)(\vec{n}) \right|^2 + \sum_{k=1}^{K} \beta_k \sum_{\vec{n}} \frac{1}{2} \left| (c_k * x)(\vec{n}) \right|^2,$$

where $c_k(\vec{n})$ denotes the impulse response of the penalty in the kth direction. By similar arguments as above, the corresponding frequency response is

$$H(\omega) = \frac{G^*(\omega)}{|G(\omega)|^2 + \sum_{k=1}^{K} \beta_k |C_k(\omega)|^2},$$

where $\omega \in \mathbb{R}^d$.

Again, consider the *denoising* case where $g(\vec{n}) = \delta[\vec{n}]$ and hence $G(\omega) = 1$. In 2D, the simplest type of penalty uses $c_1[n,m] = \delta_2[n,m] - \delta_2[n-1,m]$ and $c_2[n,m] = \delta_2[n,m] - \delta_2[n,m-1]$, so $C_k(\omega_1, \omega_2) = 1 - \mathrm{e}^{\imath\omega_k}$. Usually we use $\beta = \beta_1 = \beta_2$ also. Thus

$$\begin{aligned} H(\omega_1, \omega_2) &= \frac{1}{1 + \beta \left(\left|1 - \mathrm{e}^{\imath\omega_1}\right|^2 + \left|1 - \mathrm{e}^{\imath\omega_2}\right|^2 \right)} \\ &= \frac{1}{1 + 2\beta \left([1 - \cos\omega_1] + [1 - \cos\omega_2] \right)} \\ &\approx \frac{1}{1 + \beta \left(\omega_1^2 + \omega_2^2 \right)}, \end{aligned}$$

using the Taylor expansion $\cos x \approx 1 - x^2/2$. Thus, H is *approximately* circularly symmetric. Finding $h[n,m]$ exactly is an *open problem.*

22.13 Dynamic problems (s,srp,dynam)

Although this chapter focuses primarily on static image reconstruction, many of the general principles also apply to dynamic imaging. This section analyzes the spatial resolution properties of one family of dynamic image reconstruction methods. For simplicity we focus on the case of quadratic spatial regularization and quadratic temporal regularization [41].

Consider a dynamic model of the form

$$\boldsymbol{y}_m = \boldsymbol{A}_m \boldsymbol{x}_m + \boldsymbol{\varepsilon}_m, \quad m = 1, \dots, M,$$

where each $\boldsymbol{x}_m \in \mathbb{C}^{n_\mathrm{p}}$ and the number of columns of each $\boldsymbol{A}_m$ is n_p. (The number of rows of each $\boldsymbol{A}_m$ may differ.) Equivalently:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_M)$, $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_M)$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_M)$ and

$$\boldsymbol{A} = \mathrm{blockdiag}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_M\}.$$

Now consider a PWLS estimator of the form

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}) \tag{22.13.1}$$

$$\Psi(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2_{\boldsymbol{W}^{1/2}} + \beta \, \mathsf{R}_\mathrm{S}(\boldsymbol{x}) + \alpha \, \mathsf{R}_\mathrm{T}(\boldsymbol{x}), \tag{22.13.2}$$

where $\boldsymbol{W} = \mathrm{blockdiag}\{\boldsymbol{W}_1, \dots, \boldsymbol{W}_M\}$ and each $\boldsymbol{W}_m$ is the inverse of the covariance of $\boldsymbol{\varepsilon}_m$, at least approximately. (We assume the noise is uncorrelated between frames.) Assume the quadratic spatial regularizer has the "frame-by-frame" form

$$\mathsf{R}_\mathrm{S}(\boldsymbol{x}) = \sum_{m=1}^{M} \frac{1}{2} \|\boldsymbol{C}_{\mathrm{S}0} \boldsymbol{x}_m\|^2 \tag{22.13.3}$$

e,srp,dynam,Rsx

and the quadratic temporal regularizer has a "pixel-by-pixel" form such as

$$\mathsf{R}_\mathrm{T}(\boldsymbol{x}) = \sum_{m=2}^{M} \frac{1}{2} \|\boldsymbol{x}_m - \boldsymbol{x}_{m-1}\|^2. \tag{22.13.4}$$

e,srp,dynam,Rtx

It will simplify analysis to put both regularizers in a concise matrix form. Because the same spatial regularizer is used for each frame in (22.13.3), we can rewrite it using a *Kronecker product*:

$$\mathsf{R}_\mathrm{S}(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{C}_\mathrm{S} \boldsymbol{x}\|^2,$$

where

$$\boldsymbol{C}_\mathrm{S} \triangleq \boldsymbol{I}_M \otimes \boldsymbol{C}_{\mathrm{S}0}.$$

Because the same temporal regularizer is used for every pixel in (22.13.4), we can rewrite it as

$$\mathsf{R}_\mathrm{T}(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{C}_\mathrm{T} \boldsymbol{x}\|^2,$$

where

$$\boldsymbol{C}_\mathrm{T} = \boldsymbol{C}_{\mathrm{T}0} \otimes \boldsymbol{I}_{n_\mathrm{p}}$$

for some $M \times M$ temporal differencing matrix $\boldsymbol{C}_\mathrm{T}$.

The spatial resolution properties of the PWLS estimator (22.13.2) are given by (22.4.16):

$$\mathsf{E}[\hat{\boldsymbol{x}}] = [\mathsf{F} + \mathsf{R}]^{-1} \mathsf{F} \boldsymbol{x},$$

where the $Mn_\mathrm{p} \times Mn_\mathrm{p}$ Fisher information matrix is $\mathsf{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$, and where the Hessian of the regularizer is

$$\mathsf{R} = \beta \boldsymbol{C}_\mathrm{S}' \boldsymbol{C}_\mathrm{S} + \alpha \boldsymbol{C}_\mathrm{T}' \boldsymbol{C}_\mathrm{T} = \beta \boldsymbol{I}_M \otimes (\boldsymbol{C}_{\mathrm{S}0}' \boldsymbol{C}_{\mathrm{S}0}) + \alpha (\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}) \otimes \boldsymbol{I}_{n_\mathrm{p}}.$$

Because $\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}$ is symmetric positive semidefinite, it has a $M \times M$ matrix decomposition:

$$\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0} = \boldsymbol{Q}_\mathrm{T}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}_\mathrm{T},$$

where $\boldsymbol{\Omega} = \mathsf{diag}\{\omega_m\}$. Furthermore, because static signals are not penalized by (22.13.4), we know that $\boldsymbol{C}_{\mathrm{T}0} \mathbf{1}_M = \mathbf{0}$ and hence, without loss of generality, $\boldsymbol{Q}_\mathrm{T} \mathbf{1}_M = \boldsymbol{e}_{\mathrm{T}0} \triangleq (1, 0, \dots, 0)$ and $\omega_1 = 0$, *i.e.*, the response of the temporal regularizer to DC is zero.

22.13.1 Static system matrix

If the same system matrix (and weighting) is used for each frame repeatedly, *i.e.*, $\boldsymbol{A}_m = \boldsymbol{A}_1$ and $\boldsymbol{W}_m = \boldsymbol{W}_1$, then

$$\mathsf{F} = \boldsymbol{I}_M \otimes \mathsf{F}_1$$

$$\mathsf{F}_1 \triangleq \boldsymbol{A}_1' \boldsymbol{W}_1 \boldsymbol{A}_1$$

$$\boldsymbol{H} = \boldsymbol{I}_M \otimes \boldsymbol{H}_\mathrm{S} + \alpha (\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}) \otimes \boldsymbol{I}_{n_\mathrm{p}},$$

where

$$\boldsymbol{H}_\mathrm{S} = \mathsf{F}_1 + \beta \boldsymbol{C}_{\mathrm{S}0}' \boldsymbol{C}_{\mathrm{S}0}.$$

Thus

$$\boldsymbol{H} = \left(\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right) \left[\boldsymbol{I}_M \otimes \boldsymbol{H}_\mathrm{S} + \alpha \boldsymbol{\Omega} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right] \left(\boldsymbol{Q}_\mathrm{T} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right)$$

so

$$\boldsymbol{H}^{-1} = \left(\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right) \mathrm{blockdiag}\left\{ \left[\boldsymbol{H}_\mathrm{S} + \alpha \omega_m \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \right\}_{m=1}^{M} \left(\boldsymbol{Q}_\mathrm{T} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right).$$

Similarly,

$$\mathsf{F} = \left(\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right) \left(\boldsymbol{I}_M \otimes \mathsf{F}_1\right) \left(\boldsymbol{Q}_\mathrm{T} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right)$$

so the overall PSF matrix is

$$\boldsymbol{H}^{-1} \mathsf{F} = \left(\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right) \mathrm{blockdiag}\left\{ \left[\boldsymbol{H}_\mathrm{S} + \alpha \omega_m \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \mathsf{F}_1 \right\}_{m=1}^{M} \left(\boldsymbol{Q}_\mathrm{T} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right).$$

22.13.1.1 Static object case

To analyze the influence of the spatial regularization parameter β, it may be helpful to consider the case where the object is static, *i.e.*,

$$\boldsymbol{x} = \boldsymbol{1}_M \otimes \boldsymbol{x}_0.$$

In this case,

$$\left(\boldsymbol{Q}_\mathrm{T} \otimes \boldsymbol{I}_{n_\mathrm{p}}\right) \boldsymbol{x} = \boldsymbol{e}_{\mathrm{T}0} \otimes \boldsymbol{x}_0,$$

so, using the property $\omega_1 = 0$, the resolution properties are:

$$\begin{aligned}
\mathsf{E}[\hat{\boldsymbol{x}}] = \boldsymbol{H}^{-1} \mathsf{F} \boldsymbol{x} &= (\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}) \mathrm{blockdiag}\left\{ \left[\boldsymbol{H}_\mathrm{S} + \alpha \omega_m \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \mathsf{F}_1 \right\}_{m=1}^{M} (\boldsymbol{e}_{\mathrm{T}0} \otimes \boldsymbol{x}_0) \\
&= (\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}) \left(\boldsymbol{e}_{\mathrm{T}0} \otimes \left(\left[\boldsymbol{H}_\mathrm{S} + \alpha \omega_1 \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \mathsf{F}_1 \boldsymbol{x}_0 \right) \right) \\
&= (\boldsymbol{Q}_\mathrm{T}^{-1} \otimes \boldsymbol{I}_{n_\mathrm{p}}) \left(\boldsymbol{e}_{\mathrm{T}0} \otimes \left(\boldsymbol{H}_\mathrm{S}^{-1} \mathsf{F}_1 \boldsymbol{x}_0 \right) \right) \\
&= \boldsymbol{1} \otimes \left(\boldsymbol{H}_\mathrm{S}^{-1} \mathsf{F}_1 \boldsymbol{x}_0 \right) \\
&= \boldsymbol{1} \otimes \left(\left[\mathsf{F}_1 + \beta \boldsymbol{C}_{\mathrm{S}0}' \boldsymbol{C}_{\mathrm{S}0}\right]^{-1} \mathsf{F}_1 \boldsymbol{x}_0 \right).
\end{aligned}$$

In words, we have shown mathematically the intuitive result that if the same system model $\boldsymbol{A}_1$ is used for all frames, then the temporal regularization parameter α has no effect on the spatial resolution. This result holds for any system model, regardless of whether it is spatially shift invariant. So in this special case, the spatial regularization parameter can be set by the spatial resolution properties of a single frame, *i.e.*, by using

$$\left[\mathsf{F}_1 + \beta \boldsymbol{C}_{\mathrm{S}0}' \boldsymbol{C}_{\mathrm{S}0}\right]^{-1} \mathsf{F}_1. \tag{22.13.5}$$

e,srp,dynam,psf,static

22.13.2 General system matrices

More generally, the overall Fisher information matrix is

$$\mathsf{F} = \mathrm{blockdiag}\{\mathsf{F}_m\}_{m=1}^{M}$$

where the $n_\mathrm{p} \times n_\mathrm{p}$ Fisher information matrix for the mth frame is

$$\mathsf{F}_m = \boldsymbol{A}_m' \boldsymbol{W}_m \boldsymbol{A}_m.$$

To proceed, we assume that the per-frame Fisher information matrices and the spatial regularizer all have a common eigenspace:

$$\begin{aligned}
\mathsf{F}_m &= \boldsymbol{Q}_\mathrm{S}^{-1} \boldsymbol{\Lambda}_m \boldsymbol{Q}_\mathrm{S}, \; m = 1, \ldots, M \\
\boldsymbol{C}_{\mathrm{S}0}' \boldsymbol{C}_{\mathrm{S}0} &= \boldsymbol{Q}_\mathrm{S}^{-1} \boldsymbol{\Gamma} \boldsymbol{Q}_\mathrm{S},
\end{aligned}$$

where

$$\boldsymbol{\Lambda}_m = \mathsf{diag}\{\lambda_{m1}, \dots, \lambda_{mn_\mathrm{p}}\}.$$

This eigenspace assumption is applicable to single-coil dynamic MRI where each frame involves a subset of Cartesian k-space samples and the spatial regularizer uses periodic boundary conditions. Whether it also applies (at least approximately) to parallel MRI is an *open problem*.

Under this assumption,

$$\mathsf{F} + \beta \boldsymbol{C}_\mathrm{s}' \boldsymbol{C}_\mathrm{s} = \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}^{-1}\right) \mathrm{blockdiag}\{\boldsymbol{\Lambda}_m + \beta \boldsymbol{\Gamma}\}_{m=1}^M \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}\right)$$

so

$$\boldsymbol{H}^{-1} = \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}^{-1}\right) \left[\mathrm{blockdiag}\{\boldsymbol{\Lambda}_m + \beta \boldsymbol{\Gamma}\}_{m=1}^M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right) \otimes \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}\right)$$

and

$$\boldsymbol{H}^{-1}\mathsf{F} = \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}^{-1}\right) \left(\left[\mathrm{blockdiag}\{\boldsymbol{\Lambda}_m + \beta \boldsymbol{\Gamma}\}_{m=1}^M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right) \otimes \boldsymbol{I}_{n_\mathrm{p}}\right]^{-1} \mathrm{blockdiag}\{\boldsymbol{\Lambda}_m\}_{m=1}^M \right) \left(\boldsymbol{I}_M \otimes \boldsymbol{Q}_\mathrm{s}\right).$$

Thus far we have used a lexicographic ordering in which the spatial index varies fastest and the temporal index varies slowest. Now we consider the permutation of that ordering in which the time index varies fastest and the spatial index varies slowest. In this ordering, the PSF becomes

$$\tilde{\boldsymbol{H}}^{-1}\tilde{\mathsf{F}} = \left(\boldsymbol{Q}_\mathrm{s}^{-1} \otimes \boldsymbol{I}_M\right) \left(\left[\mathrm{blockdiag}\{\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M\}_{j=1}^{n_\mathrm{p}} + \alpha \boldsymbol{I}_{n_\mathrm{p}} \otimes \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \mathrm{blockdiag}\{\boldsymbol{D}_j\}_{j=1}^{n_\mathrm{p}} \right) \left(\boldsymbol{Q}_\mathrm{s} \otimes \boldsymbol{I}_M\right)$$

$$= \left(\boldsymbol{Q}_\mathrm{s}^{-1} \otimes \boldsymbol{I}_M\right) \mathrm{blockdiag}\left\{ \left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j \right\}_{j=1}^{n_\mathrm{p}} \left(\boldsymbol{Q}_\mathrm{s} \otimes \boldsymbol{I}_M\right),$$

where each $M \times M$ matrix $\boldsymbol{D}_j$ is defined by

$$\boldsymbol{D}_j = \mathsf{diag}\{\lambda_{1j}, \dots, \lambda_{Mj}\}.$$

In words, the expression

$$\left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j$$

corresponds to temporal smoothing of the jth frequency component.

22.13.2.1 Static object case

When the object is static, then in the "time fastest" lexicographic ordering:

$$\tilde{\boldsymbol{x}} = \boldsymbol{x}_0 \otimes \boldsymbol{1}_M.$$

In this case,

$$\left(\boldsymbol{Q}_\mathrm{s} \otimes \boldsymbol{I}_M\right) \tilde{\boldsymbol{x}} = \left(\boldsymbol{Q}_\mathrm{s} \boldsymbol{x}_0\right) \otimes \boldsymbol{1}_M.$$

Let $\boldsymbol{X}_0 = \boldsymbol{Q}_\mathrm{s} \boldsymbol{x}_0 = (X_1, \dots, X_{n_\mathrm{p}})$ denote the spatial transform of $\boldsymbol{x}_0$. Then

$$\begin{aligned} \tilde{\boldsymbol{H}}^{-1}\tilde{\mathsf{F}}\tilde{\boldsymbol{x}} &= \left(\boldsymbol{Q}_\mathrm{s}^{-1} \otimes \boldsymbol{I}_M\right) \mathrm{blockdiag}\left\{ \left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j \right\}_{j=1}^{n_\mathrm{p}} \left(\boldsymbol{X}_0 \otimes \boldsymbol{1}_M\right) \\ &= \left(\boldsymbol{Q}_\mathrm{s}^{-1} \otimes \boldsymbol{I}_M\right) \begin{bmatrix} \vdots \\ X_j \left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j \boldsymbol{1}_M \\ \vdots \end{bmatrix}. \end{aligned}$$

Even though the object $\tilde{\boldsymbol{x}}$ is static here, in general the mean reconstruction given above will have time-varying frequency content due to changes in the system matrix between frames. For the purpose of choosing the spatial regularization parameter it may be helpful to examine the *time average* of the mean reconstructed image:

$$\begin{aligned} \left(\boldsymbol{I}_{n_\mathrm{p}} \otimes \frac{1}{M}\boldsymbol{1}_M'\right) \tilde{\boldsymbol{H}}^{-1}\tilde{\mathsf{F}}\tilde{\boldsymbol{x}} &= \left(\boldsymbol{Q}_\mathrm{s}^{-1} \otimes \frac{1}{M}\boldsymbol{1}_M'\right) \begin{bmatrix} \vdots \\ X_j \left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j \boldsymbol{1}_M \\ \vdots \end{bmatrix} \\ &= \boldsymbol{Q}_\mathrm{s}^{-1} \begin{bmatrix} \vdots \\ X_j \frac{1}{M}\boldsymbol{1}_M' \left[\boldsymbol{D}_j + \beta\gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\mathrm{T}0}' \boldsymbol{C}_{\mathrm{T}0}\right)\right]^{-1} \boldsymbol{D}_j \boldsymbol{1}_M \\ \vdots \end{bmatrix} \end{aligned}$$

$$= \boldsymbol{Q}_{\text{S}}^{-1} \boldsymbol{L}(\beta, \alpha) \boldsymbol{X}_0,$$

where $\boldsymbol{L}$ is a $n_{\text{p}} \times n_{\text{p}}$ diagonal "frequency response" matrix with elements

$$L_{jj}(\beta, \alpha) = \frac{1}{M} \mathbf{1}_M' \left[\boldsymbol{D}_j + \beta \gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\text{T0}}' \boldsymbol{C}_{\text{T0}} \right) \right]^{-1} \boldsymbol{D}_j \mathbf{1}_M.$$

In the case of a static system matrix, $\boldsymbol{D}_j = \lambda_{1j} \boldsymbol{I}_M$, and $L_{jj}(\beta, \alpha) = \frac{\lambda_{1j}}{\lambda_{1j} + \beta \gamma_j}$ which is independent of α and concurs with (22.13.5).

Consider the case where $\alpha = 0$:

$$L_{jj}(\beta, 0) = \frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_{mj}}{\lambda_{mj} + \beta \gamma_j}.$$

Now assume that the jth frequency component is "sampled" $0 \leq M_j \leq M$ times, and suppose that for each such sample we have the same eigenvalue: $\lambda_{mj} = \lambda_j$. (This is reasonable for single-coil MRI but investigating it for parallel MRI is an *open problem*.) Then the frequency response elements simplify to

$$L_{jj}(\beta, 0) = \frac{M_j}{M} \frac{\lambda_j}{\lambda_j + \beta \gamma_j}.$$

Compared to the static system matrix case, the jth frequency component is reduced by M_j / M when α is very small. This does not change the spatial resolution, but simply attenuates the values of the (time averaged) mean reconstructed image.

Now consider the limit as $\alpha \to \infty$. By derivations analogous to (2.6.2):

$$\lim_{\alpha \to \infty} \left[\boldsymbol{D}_j + \beta \gamma_j \boldsymbol{I}_M + \alpha \left(\boldsymbol{C}_{\text{T0}}' \boldsymbol{C}_{\text{T0}} \right) \right]^{-1} = \mathbf{1}_M (\mathbf{1}_M' \left(\boldsymbol{D}_j + \beta \gamma_j \boldsymbol{I}_M \right) \mathbf{1}_M)^{-1} \mathbf{1}_M',$$

because $\boldsymbol{C}_{\text{T0}}$ in (22.13.4) uses finite differences in time. So

$$\lim_{\alpha \to \infty} L_{jj}(\beta, \alpha) = \frac{1}{\mathbf{1}_M' \left(\boldsymbol{D}_j + \beta \gamma_j \boldsymbol{I}_M \right) \mathbf{1}_M} \mathbf{1}_M' \boldsymbol{D}_j \mathbf{1}_M = \frac{\sum_{m=1}^{M} \lambda_{mj}}{\sum_{m=1}^{M} \left(\lambda_{mj} + \beta \gamma_j \right)}.$$

If the jth frequency component is sampled M_j out of M times, then

$$\lim_{\alpha \to \infty} L_{jj}(\beta, \alpha) = \frac{M_j \lambda_{1j}}{M_j \lambda_{1j} + M \beta \gamma_j} = \frac{\lambda_{1j}}{\lambda_{1j} + \frac{M}{M_j} \beta \gamma_j}.$$

In other words, for large values of α, the *effective spatial regularization parameter* is *increased* by a factor M / M_j relative to the static case.

s,srp,note

22.14 Notes (s,srp,note)

Except for the continuous-discrete analysis in §22.2.2, this chapter focused on reconstruction problems where there is no system model mismatch, *i.e.*, we assume we have the correct system model $\boldsymbol{A}$ in $\bar{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$. One can generalize the analysis to consider effects of model mismatch [42].

In statistics literature, often regularization parameters are chosen to minimize expected squared error as the number of data points increases asymptotically, *e.g.*, [43].

Use local Fourier analytical approximation to F_k and R_k to help choose β via Φ_{GCV} in (2.5.37)?

IRT *See* `tomo2_beta_test`.

To maintain consistent spatial resolution (in mm) as one varies n_φ , $\triangle_{\text{R}}$, FOV, or $\triangle_{\text{X}}$, one must multiply β by $n_\varphi / \triangle_{\text{R}}$. (For consistent resolution in pixels, one would also multiply by $\triangle_{\text{X}}^3$.) This follows from analyzing how the diagonal elements of $\boldsymbol{A}'\boldsymbol{A}$ vary with those scan parameters.

[47] similar considerations used for partial volume correction

[48] trade-off between confocal pinhole size and amount of deconvolution needed

[49] used quadratic penalty based on system's SVD to make uniform variance.

[50] Generalized matrix inverse reconstruction for SPECT using a weighted singular value spectrum @an uses apodized SVD rather than truncated to get better PSFs. essentially local impulse response "design" subject to variance constraint!

[15] uniform contrast recovery coefficient (CRC). iterative coordinate descent to optimize a modified quadratic penalty function.

[51, 52] oriented approach

22.15 Problems (s,srp,prob)

s,srp,prob

p,srp,four,cart

Problem 22.1 *Analysis the local impulse response of a PWLS estimator for the case: white noise, $\boldsymbol{W} = \boldsymbol{I}$; Cartesian Fourier samples, $\boldsymbol{A} = \boldsymbol{Q}$; and Tikhonov regularization, $\mathsf{R}(\boldsymbol{x}) = \beta \frac{1}{2} \|\boldsymbol{x}\|^2$.*

p,srp,eql

Problem 22.2 *For a Poisson emission tomography problem with a penalized-likelihood estimator, compute the local impulse response using the perturbation approximation (22.5.1), the system of equations (22.5.2), and the DFT approximation (22.7.3) and compare. (Need typed.)*

p,srp,step

Problem 22.3 *For linear shift invariant systems, there is a simple relationship between a system's impulse response and its step response. In the context of imaging problems, "step response" could be important in terms of analyzing how well edges are preserved by a given reconstruction method.*
Propose a criterion for the step response of nonlinear image reconstruction methods and analyze it for penalized-likelihood estimators with edge preserving regularization. Hint: see [32]. The 1D case for a Huber potential function was considered in [53]. (Solve?)

p,srp,fish,kapj,1

Problem 22.4 *Compare (22.9.2) with the approximation (22.9.6) for a single projection view in the limit as the radial sampling approaches zero.*

(Solve?)

Problem 22.5 *Consider a reconstruction problem with noiseless data under the linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and convex, differentiable, edge-preserving regularization. Examine center point $\hat{x}_j$ of the local impulse response $\frac{\hat{\boldsymbol{x}}(\boldsymbol{y}+\varepsilon \boldsymbol{A}\boldsymbol{e}_j) - \hat{\boldsymbol{x}}(\boldsymbol{y})}{\varepsilon}$ as a function of ε varying between nearly zero and much larger values. For small ε values the plot should be nearly flat, but as ε increases there may be a knee in the curve. Analyze this problem to determine the approximate location of the knee and find the slope for large ε, (assuming the curve is approximately affine beyond the knee). (Solve?)*

22.16 Bibliography

rayleigh:1874:otm
[1] L. Rayleigh. On the manufacture and theory of diffraction gratings. *Philos. Mag.*, 47:81–93, 1874. 22.2

unser:87:anr
[2] M. Unser, B. L. Trus, and A. C. Steven. A new resolution criterion based on spectral signal to noise ratio. *Ultramicroscopy*, 23(1):39–51, 1987. 22.2

kosarev:90:ssl
[3] E. L. Kosarev. Shannon's superresolution limit for signal recovery. *Inverse Prob.*, 6(1):55–76, February 1990. Correction in 6(3):479, June 1990. 22.2

dendekker:97:ras
wang:99:aaf
[4] A. J. den Dekker and A. van den Bos. Resolution: a survey. *J. Opt. Soc. Am. A*, 14(3):547–57, March 1997. 22.2

[5] G. Wang and Y. Li. Axiomatic approach for quantification of image resolution. *IEEE Signal Proc. Letters*, 6(10):257–8, October 1999.

difrancia:55:rpa
helstrom:64:tda
[6] G. T. di Francia. Resolving power and information. *J. Opt. Soc. Am.*, 45(7):497–501, July 1955. 22.2

[7] C. W. Helstrom. The detection and resolution of optical signals. *IEEE Trans. Info. Theory*, 10(4):275–87, October 1964. 22.2

lucy:92:rlf
[8] L. B. Lucy. Resolution limits for deconvolved images. *The Astronomical Journal*, 104(3):1260–5, September 1992. 22.2

vandenbos:92:ura
[9] A. van den Bos. Ultimate resolution: a mathematical framework. *Ultramicroscopy*, 47(1-3):298–306, November 1992. 22.2

dharanipragada:96:rli
[10] S. Dharanipragada and K. S. Arun. Resolution limits in signal recovery. *IEEE Trans. Sig. Proc.*, 44(3):546–61, March 1996. 22.2

shahram:04:ibt
[11] M. Shahram and P. Milanfar. Imaging below the diffraction limit: a statistical analysis. *IEEE Trans. Im. Proc.*, 13(5):677–89, May 2004.

williamson:72
[12] R. E. Williamson, R. H. Crowell, and H. F. Trotter. *Calculus of vector functions*. Prentice-Hall, New Jersey, 1972. 22.4

elbakri:03:eaa
[13] I. A. Elbakri and J. A. Fessler. Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography. In *Proc. SPIE 5032 Medical Imaging 2003: Image Proc.*, pages 1839–50, 2003. 22.6

fessler:96:mav
[14] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996. 22.8, 22.9, 22.10, 22.13

qi:00:ran
[15] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3D PET. *IEEE Trans. Med. Imag.*, 19(5):493–506, May 2000. 22.8, 22.10, 22.13, 22.20

li:04:aeo
[16] Q. Li, E. Asma, J. Qi, J. R. Bading, and R. M. Leahy. Accurate estimation of the Fisher information matrix for the PET image reconstruction problem. *IEEE Trans. Med. Imag.*, 23(9):1057–64, September 2004. 22.8

hero:96:eeb
[17] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform CR bound. *IEEE Trans. Sig. Proc.*, 44(8):2026–41, August 1996. 22.9

chan:88:aoc
[18] T. F. Chan. An optimal circulant preconditioner for Toeplitz systems. *SIAM J. Sci. Stat. Comp.*, 9(4):766–71, July 1988. 22.10

tyrtyshnikov:92:oas
[19] E. Tyrtyshnikov. Optimal and superoptimal circulant preconditioners. *SIAM J. Matrix. Anal. Appl.*, 13(2):459–73, 1992. 22.10

chan:07
[20] R. H-F. Chan and X-Q. Jin. *An introduction to iterative Toeplitz solvers*. Soc. Indust. Appl. Math., Philadelphia, 2007.

chan:96:cgm
[21] R. H. Chan and M. K. Ng. Conjugate gradient methods for Toeplitz systems. *SIAM Review*, 38(3):427–82, September 1996. 22.10

clinthorne:93:pmf
[22] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imag.*, 12(1):78–83, March 1993. 22.10

fessler:99:cgp
[23] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans. Im. Proc.*, 8(5):688–99, May 1999. 22.10

stayman:04:cfn
[24] J. W. Stayman and J. A. Fessler. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. *IEEE Trans. Med. Imag.*, 23(3):269–84, March 2004. 22.10, 22.13, 22.15

stayman:04:eco
[25] J. W. Stayman and J. A. Fessler. Efficient calculation of resolution and covariance for fully-3D SPECT. *IEEE Trans. Med. Imag.*, 23(12):1543–56, December 2004. 22.10, 22.13

olafsson:06:sra
[26] V. Olafsson, J. A. Fessler, and D. C. Noll. Spatial resolution analysis of iterative image reconstruction with separate regularization of real and imaginary parts. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 5–8, 2006. 22.11

chang:78:amf
[27] L. T. Chang. A method for attenuation correction in radionuclide computed tomography. *IEEE Trans. Nuc. Sci.*, 25(1):638–643, February 1978. 22.11

fessler:96:srp
[28] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996. 22.12, 22.15

zubal:90:hra
[29] G. Zubal, G. Gindi, M. Lee, C. Harrell, and E. Smith. High resolution anthropomorphic phantom for Monte Carlo analysis of internal radiation sources. In *IEEE Symposium on Computer-Based Medical Systems*, pages 540–7, 1990. 22.12

stayman:00:rfu
[30] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–15, June 2000. 22.12, 22.13

shi:09:qrd
[31] H. R. Shi and J. A. Fessler. Quadratic regularization design for 2D CT. *IEEE Trans. Med. Imag.*, 28(5):645–56, May 2009. 22.13, 22.14

ahn:06:srp
[32] S. Ahn and R. M. Leahy. Spatial resolution properties of nonquadratically regularized image reconstruction for PET. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 287–90, 2006. 22.13, 22.21

ahn:08:aor
[33] S. Ahn and R. M. Leahy. Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET. *IEEE Trans. Med. Imag.*, 27(3):413–24, March 2008. 22.13

fessler:94:uqp
[34] J. A. Fessler and W. L. Rogers. Uniform quadratic penalties cause nonuniform image resolution (and sometimes vice versa). In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 1915–9, 1994. 22.13

fessler:03:aat
[35] J. A. Fessler. Analytical approach to regularization design for isotropic spatial resolution. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 2022–6, 2003. 22.14

wang:11:sfm
[36] J. Wang, K. Sauer, J-B. Thibault, Z. Yu, and C. Bouman. Spectrally focused Markov random field image modeling in 3D CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 152–5, 2011. 22.14

mustafovic:04:odo
[37] S. Mustafovic and K. Thielemans. Object dependency of resolution in reconstruction algorithms with inter-iteration filtering applied to PET data. *IEEE Trans. Med. Imag.*, 23(4):433–46, April 2004. 22.14

silverman:90:ase
[38] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Royal Stat. Soc. Ser. B*, 52(2):271–324, 1990. 22.15

ahn:03:gci
[39] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.*, 22(5):613–26, May 2003. 22.15

unser:91:rrf
[40] M. Unser, A. Aldroubi, and M. Eden. Recursive regularization filters: design, properties, and applications. *IEEE Trans. Patt. Anal. Mach. Int.*, 13(3):272–7, March 1991. 22.15

khalsa:07:rpi
[41] K. A. Khalsa and J. A. Fessler. Resolution properties in regularized dynamic MRI reconstruction. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 456–9, 2007. 22.17

fessler:97:srp
[42] J. A. Fessler. Spatial resolution properties of penalized weighted least-squares image reconstruction with model mismatch. Technical Report 308, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, March 1997. Available from `web.eecs.umich.edu/~fessler`. 22.20

kohler:01:nre
[43] M. Kohler and A. Krzyzak. Nonparametric regression estimation using penalized least squares. *IEEE Trans. Info. Theory*, 47(7):3054–8, November 2001. 22.20

thompson:91:aso
[44] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington. A study of methods or choosing the smoothing parameter in image restoration by regularization. *IEEE Trans. Patt. Anal. Mach. Int.*, 13(4):326–39, April 1991.

hutchinson:90:ase
[45] M. F. Hutchinson. A stochastic estimator for the trace of the influence matrix for Laplacian smoothing splines. *Comm. in Statistics - Simulation and Computation*, 19(2):433–50, 1990.

golub:97:gcv
[46] G. H. Golub and U. von Matt. Generalized cross-validation for large-scale problems. *J. Computational and Graphical Stat.*, 6(1):1–34, March 1997.

boening:04:sor
[47] G. Boening, P. H. Pretorius, and M. A. King. Study of relative quantification of Tc-99m with partial volume effect and spillover correction for SPECT oncology imaging. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 5, pages 2705–9, 2004. 22.20

conchello:94:etd
[48] J-A. Conchello, J. J. Kim, and E. W. Hansen. Enhanced three-dimensional reconstruction from confocal scanning microscope images. II. Depth discrimination versus signal-to-noise ratio in partically confocal images. *Appl. Optics*, 33(17):3740–50, June 1994. 22.20

cohenbacrie:97:rri
[49] C. Cohen-Bacrie, Y. Goussard, and R. Guardo. Regularized reconstruction in electrical impedance tomography using a variance uniformization constraint. *IEEE Trans. Med. Imag.*, 16(5):562–71, October 1997. 22.20

smith:96:gmi
[50] M. F. Smith. Generalized matrix inverse reconstruction for SPECT using a weighted singular value spectrum. *IEEE Trans. Nuc. Sci.*, 43(3-2):2008–17, June 1996. 22.20

nuyts:02:cbp
[51] J. Nuyts and J. A. Fessler. Comparison between penalized-likelihood image reconstruction for uniform spatial resolution and post-smoothed ML-EM. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 895–9, 2002. 22.20

nuyts:03:apl
[52] J. Nuyts and J. A. Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to post-smoothed maximum-likelihood with matched spatial resolution. *IEEE Trans. Med. Imag.*, 22(9):1042–52, September 2003. 22.20

li:98:cfs
[53] S. Z. Li. Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing. *IEEE Trans. Patt. Anal. Mach. Int.*, 20(9):916–32, September 1998. 22.21