Image Processing 2, Exercise 8

1 Classical image reconstruction

[basic] This exercise will teach you how to design a "classical" image reconstruction algorithm. Given some physical model, it will take you through all the design steps from the discretization to the specification of an iterative reconstruction scheme.

The term "classical" refers to the reconstruction algorithms that can be derived from the minimization of a quadratic functional. The idea is that such functionals have a unique minimum, which can usually be determined in closed form. This then yields a linear reconstruction which often has a direct matrix-free implementation. Otherwise, the reconstruction can be performed iteratively by steepest descent or some faster variant (e.g., conjugate gradient).

The discretization step is a prerequisite for computational imaging. In practice, one always start by designing a classical algorithm, which is easier to debug and which encompasses all the knowledge of the physics. In a later stage, one can refine the reconstruction by incorporating additional constraints such as positivity, some improved form of regularization, and/or even the appropriate deployment of neural networks.

The continuous-domain modelling of our imaging system (1D for simplicity) in the noise-free scenario is $g_0(x) = (h * s)(x)$ where $h(x) = \mathrm{rect}(x/3)$. To reconstruct the signal, we propose to represent the signal in a "pixelated" basis as $s(x) = \sum_{k=1}^{K} s_k \beta^0(x-k)$ where $\beta^0(x) = \mathrm{rect}(x)$.

Given the noisy measurements $\mathbf{y} = (y_m) \in \mathbb{R}^M$ with $y_m = g_0(m) + n_m$, the task is to reconstruct the unknown signal coefficients $\mathbf{s} = (s_k) \in \mathbb{R}^K$ such as to minimise a suitable cost functional.

(a) Discretization: For simplicity, we take $K = M$. Specify the system matrix $\mathbf{H} \in \mathbb{R}^{M\times M}$ such that the forward imaging model can be written as $\mathbf{y} = \mathbf{Hs} + \mathbf{n}$ where $\mathbf{n} = (n_m)$ represents the measurement noise.

Solution: The transcription of the physical convolutional model for the mth noise-free measurement is $g_0(m) = \int_{\mathbb{R}} s(x)h(m-x)\mathrm{d}x = \langle s, h(m-\cdot)\rangle = \sum_{k=1}^{M} s_k \langle \beta^0(\cdot - k), h(m-\cdot)\rangle$. Consequently, the entries of the system matrix are given by

$$\mathbf{H}_{m,k} = \langle \beta^0(\cdot - k), h(m-\cdot)\rangle = \begin{cases} 1, & \text{if } k = m-1 \ \& \ k \geq 1 \\ 1, & k = m \\ 1, & k = m+1 \ \& \ k \leq M \\ 0, & \text{otherwise} \end{cases}$$

which amounts to a tridiagonal matrix of ones. Since both β^0 and $h = \beta^0(\cdot+1) + \beta^0 + \beta^0(\cdot-1)$ are indicator functions, the inner product of their shifted versions is either one—when the functions are overlapping—or zero otherwise. Another ways to put it is that $\mathbf{H}$ is the Toeplitz matrix associated with the mask $(1,1,1)$, which corresponds to the weights of convolution kernel h in the (orthonormal) B-spline basis $\{\beta^0(\cdot - k)\}$.

(b) We now formulate the reconstruction problem as the minimization of the cost functional

$$J(\mathbf{y}, \mathbf{s}) = \|\mathbf{\Lambda}(\mathbf{y} - \mathbf{Hs})\|^2 + \lambda\|\mathbf{s} - \mathbf{s}_0\|^2, \tag{1}$$

where $\mathbf{s}_0$ is a known reference signal, $\lambda \geq 0$ a tuning parameter and $\mathbf{\Lambda} = \mathrm{diag}(a_1, \dots, a_M)$ is a diagonal matrix of weights $a_m > 0$.

Motivation: The use of this particular cost functional is our way of injecting prior information. Specifically, we want the reconstruction to be "reasonably" close to $\mathbf{s}_0$, while we are also weighting the contribution of a measurement in inverse proportion to the standard deviation of the noise which may be detector-dependent; e.g., $a_m \propto 1/\sigma_m$.

Your first task is to calculate the gradient of $J(\mathbf{y}, \mathbf{s})$ with respect to $\mathbf{s}$ (the unknown).

Solution: First, we expand (1) and put it in a form that facilitates the use of our vector calculus:

$$\begin{aligned} J(\mathbf{y},\mathbf{s}) &= \|\mathbf{\Lambda y}\|^2 + \mathbf{s}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H}\mathbf{s} - 2\mathbf{y}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H}\mathbf{s} + \lambda\mathbf{s}^\mathsf{T}\mathbf{s} - 2\lambda\mathbf{s}_0^\mathsf{T}\mathbf{s} + \|\mathbf{s}_0\|^2 \\ &= \mathbf{s}^\mathsf{T}(\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{I})\mathbf{s} - 2(\mathbf{y}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{s}_0^\mathsf{T})\mathbf{s} + \|\mathbf{\Lambda y}\|^2 + \|\mathbf{s}_0\|^2 \\ &= \mathbf{s}^\mathsf{T}\mathbf{A}\mathbf{s} - 2\mathbf{a}^\mathsf{T}\mathbf{s} + a_0, \end{aligned}$$

where the matrix $\mathbf{A} = \mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{I} \in \mathbb{R}^{M\times M}$ is symmetric, the vector $\mathbf{a} = \mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{y} + \lambda\mathbf{s}_0 \in \mathbb{R}^M$ is fixed, and $a_0 = \|\mathbf{\Lambda y}\|^2 + \|\mathbf{s}_0\|^2$ is a constant. With the help of the identities in Slide 10-27 (vector calculus), we then get

$$\frac{\partial J(\mathbf{y},\mathbf{s})}{\partial \mathbf{s}} = 2\mathbf{A}\mathbf{s} - 2\mathbf{a} = 2(\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{I})\mathbf{s} - 2(\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{y} + \lambda\mathbf{s}_0).$$

(c) Give the closed form of the signal reconstruction: $\arg\min_{\mathbf{s}\in\mathbb{R}^M} J(\mathbf{y},\mathbf{s})$.

Solution: All quadratic forms are strictly convex and so is our loss function $J(\mathbf{y},\mathbf{s})$. Its unique minimiser is found be setting $\frac{\partial J(\mathbf{y},\mathbf{s})}{\partial \mathbf{s}} = \mathbf{0} \Leftrightarrow 2\mathbf{A}\mathbf{s} = 2\mathbf{a}$, which yields

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{a} = (\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{I})^{-1}(\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{y} + \lambda\mathbf{s}_0). \quad (2)$$

We observe that the underlying matrix $\mathbf{A}$ is (strictly) positive-definite—and hence invertible—because of the presence of the diagonal term $\lambda\mathbf{I}$. As cross-check, we can also verify that (2) is compatible with the Tikhonov solution $\mathbf{s} = (\mathbf{H}^\mathsf{T}\mathbf{H} + \lambda\mathbf{L}^\mathsf{T}\mathbf{L})^{-1}\mathbf{H}^\mathsf{T}\mathbf{y}$ given in the course if we set $\mathbf{s}_0 = \mathbf{0}$, $\mathbf{\Lambda} = \mathbf{I}$ and select the regularization operator $\mathbf{L} = \mathbf{I}$ (identity).

(d) (i) Give a short description of an iterative gradient-based reconstruction algorithm with initial condition $\mathbf{s}^{(0)} = \mathbf{s}_0$.
(ii) Modify your algorithm in order to impose the positivity of your reconstruction.

Solution: (i) The reconstruction algorithm is given as pseudo-code. The only parameter is the step size γ which needs to be selected sufficiently small for the algorithm to converge.

Initialisation:
$k \leftarrow 0$; $\mathbf{s}^{(0)} \leftarrow \mathbf{s}_0$;
$\mathbf{u}_0 = \mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{y} + \lambda\mathbf{s}_0$;
Do
1: $\mathbf{s}^{(k+1)} \leftarrow \mathbf{s}^{(k)} + \gamma\, 2\big(\mathbf{u}_0 - (\mathbf{H}^\mathsf{T}\mathbf{\Lambda}^2\mathbf{H} + \lambda\mathbf{I})\mathbf{s}^k\big)$
2: $k \leftarrow k+1$
3:
Until stop criterion

(ii) To enforce positivity, one simply inserts in the loop an additional pointwise projection, where all non-negative entries of $\mathbf{s}^{(k)}$ are clipped to zero:
3: $\mathbf{s}^{(k)} \leftarrow \mathrm{ReLU}(\mathbf{s}^{(k)})$

The latter does not affect convergence because the projection on the set of nonnegative vectors is a nonexpansive map.

2 Convex functionals

Most image reconstruction algorithms are derived from the minimization of a convex functional. You should therefore be able to recognize and manipulate such functionals.

Motivation: The convexity property ensures that the standard iterative procedures (based on steepest descent and proximal updates) converge to the solution.

We recall that a functional $f : \mathbb{R}^d \to \mathbb{R}$ is convex if, for all $\lambda \in [0, 1]$ and all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \tag{3}$$

Under the assumption that the functions $\phi, \psi : \mathbb{R} \to \mathbb{R}$ are convex, prove the following properties.

(a) $a_0\phi(x)$ is convex for any fixed $a_0 \in \mathbb{R}^+$.

Solution: It suffices to check that (3) holds. Specifically, for all $x, y \in \mathbb{R}$,

$$\begin{aligned} a_0\phi\big(\lambda x + (1-\lambda)y\big) &\leq a_0\big(\lambda\phi(x) + (1-\lambda)\phi(y)\big) \qquad \text{(using (3) with } d = 1) \\ &\leq \lambda a_0\phi(x) + (1-\lambda)a_0\phi(y), \end{aligned}$$

which is the desired inequality.
If ϕ is twice differentiable, an even simpler argument is $\phi''(x) \geq 0 \Rightarrow a_0\phi''(x) \geq 0$ for all $x \in \mathbb{R}$.

(b) $g(x) = \phi(x) + \psi(x)$ is convex.

Solution: Indeed,

$$\begin{aligned} \phi\big(\lambda x + (1-\lambda)y\big) + \psi\big(\lambda x + (1-\lambda)y\big) &\leq \lambda\phi(x) + (1-\lambda)\phi(y) + \lambda\psi(x) + (1-\lambda)\psi(y) \\ &\qquad \text{(using (3) twice)} \\ &\leq \lambda\big(\phi(x) + \psi(x)\big) + (1-\lambda)\big(\phi(y) + \psi(y)\big) \end{aligned}$$

(c) The function $x \mapsto \phi(x) + a_2x^2 + a_1x + a_0$ with $a_2, a_1 \in \mathbb{R}^+$ and $a_0 \in \mathbb{R}$ is convex.

Solution: It suffices to observe (or prove) that the functions x^2 and x are convex and to then deduce the result from Properties (a) and (b).

(d) The function $g : \mathbb{R}^d \to \mathbb{R}$ where $g(\mathbf{x}) = \phi(\mathbf{w}^\mathsf{T}\mathbf{x} + b_0)$ with parameters $\mathbf{w} \in \mathbb{R}^d$ and $b_0 \in \mathbb{R}$ is convex.

Solution: Let us introduce the auxiliary variables $t_1 = \mathbf{w}^\mathsf{T}\mathbf{x} + b_0 \in \mathbb{R}$ and $t_2 = \mathbf{w}^\mathsf{T}\mathbf{x} + b_0 \in \mathbb{R}$. The convexity of ϕ then gives

$$\begin{aligned} g\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) &= \phi\big(\mathbf{w}^\mathsf{T}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + b_0\big) = \phi\big(\lambda t_1 + (1-\lambda)t_2\big) \\ &\qquad \text{(using the fact that } b_0 = \lambda b_0 + (1-\lambda)b_0) \\ &\leq \lambda\phi(t_1) + (1-\lambda)\phi(t_2) = \lambda\phi\big(\mathbf{w}^\mathsf{T}\mathbf{x} + b_0) + (1-\lambda)\phi\big(\mathbf{w}^\mathsf{T}\mathbf{x} + b_0) \\ &\leq \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}) \end{aligned}$$

(e) Variation of a theme. Let us consider a convex function $\Phi : \mathbb{R}^d \to \mathbb{R}$ and an affine map $\mathbb{R}^{d_0} \to \mathbb{R}^d : \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{d \times d_0}$ and $\mathbf{b} \in \mathbb{R}^d$ are some fixed parameters. Prove that the composed function $g : \mathbb{R}^{d_0} \to \mathbb{R}$ with $g(\mathbf{x}) = \Phi(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex .

Solution: For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d_0}$, we identify the auxiliary variables $\mathbf{t}_1 = \mathbf{A}\mathbf{x} + \mathbf{b} \in \mathbb{R}^d$ and $\mathbf{t}_2 = \mathbf{A}\mathbf{x} + \mathbf{b} \in \mathbb{R}^d$. The convexity of Φ then gives

$$\begin{aligned} g\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) &= \Phi\Big(\mathbf{A}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + \mathbf{b}\Big) = \Phi\big(\lambda\mathbf{t}_1 + (1-\lambda)\mathbf{t}_2\big) \\ &\qquad \text{(using the fact that } \mathbf{b} = \lambda\mathbf{b} + (1-\lambda)\mathbf{b}) \\ &\leq \lambda\Phi(\mathbf{t}_1) + (1-\lambda)\Phi(\mathbf{t}_2) = \lambda\Phi\big(\mathbf{A}\mathbf{x} + \mathbf{b}\big) + (1-\lambda)\Phi\big(\mathbf{A}\mathbf{x} + \mathbf{b}\big) \\ &\leq \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}), \end{aligned}$$

which is the desired inequality.

3 Proximal Operators

[basic] Proximal operators are extremely useful components for the design of efficient iterative optimization/reconstruction algorithms.

We recall that the proximal operator of a convex and l.s.c. function $f : \mathbb{R}^d \to \mathbb{R}$ is given by

$$\mathrm{prox}_f(\mathbf{x}) = \arg\min_{\mathbf{z}\in\mathbb{R}^d} \tfrac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + f(\mathbf{z}). \tag{4}$$

(a) Scalar scenario. Let $g(x) = \lambda|x|^3$ where $\lambda > 0$ is a fixed constant. First, show that g is convex, and, then determine its proximal operator $\mathrm{prox}_g(x)$.

Solution: The first and second derivatives of g are $g'(x) = 3\lambda x^2 \mathrm{sign}(x)$ and $g''(x) = 3\lambda|x|$, respectively. Since $g''(x) \geq 0$ for all $x \in \mathbb{R}$, we readily deduce that g is convex.

The proximal operator $\mathrm{prox}_g(x)$ is the minimizer of $J(z,x) = \frac{1}{2}(z-x)^2 + \lambda|z|^3$ over $z \in \mathbb{R}$ with x fixed. To find this minimum, we differentiate with respect to z and set the derivative to zero, which yields the optimality condition $\frac{\partial}{\partial z}J(z,x) = (z-x) + 3\lambda z^2 \mathrm{sign}(z) = 0 \Leftrightarrow x = g'(z) + z$. In order to express z as a function of x, we first assume that $z \geq 0$, which results in the quadratic equation $3\lambda z^2 + z - x = 0$ whose roots are $z_{1,2} = \frac{-1\pm\sqrt{1+12\lambda x}}{6\lambda}$, with only the first one being real-valued positive (consistency with our hypothesis) provided that $x \geq 0$. A similar reasoning applies for $z < 0$, with the roots now being $z_{1,2} = \frac{-1\pm\sqrt{1-12\lambda x}}{-6\lambda}$. Ultimately, we get

$$\mathrm{prox}_g(x) = \begin{cases} \frac{-1+\sqrt{1+12\lambda x}}{6\lambda}, x \geq 0 \\ \frac{1-\sqrt{1-12\lambda x}}{6\lambda}, x < 0 \end{cases} = \mathrm{sign}(x)\frac{-1+\sqrt{1+12\lambda|x|}}{6\lambda}, \tag{5}$$

whose graph is shown in Figure 1.

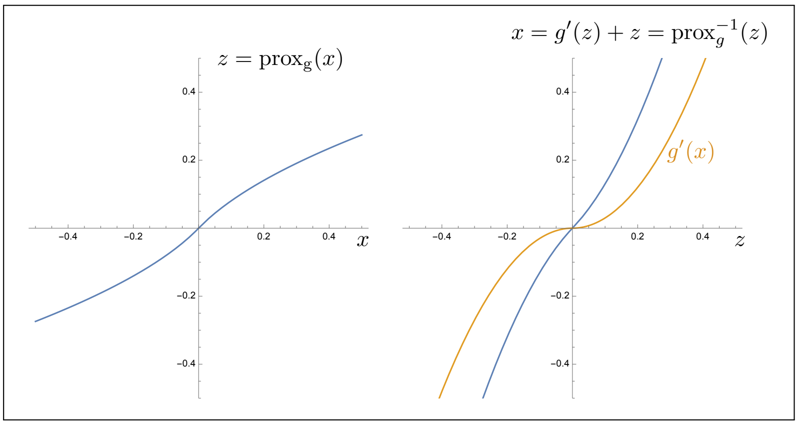

Fig 1: Proximal operator of $g(x) = |x|^3$ and related functions.

(b) Quadratic regularization functional. Your task is to derive the explicit form of the proximal operator of the functional $f_2 : \mathbb{R}^d \to \mathbb{R}$ with $f_2(\mathbf{x}) = \frac{\lambda}{2}\|\mathbf{L}\mathbf{x}\|_2^2$ where $\mathbf{L} \in \mathbb{R}^{d\times d}$ is a square matrix and $\lambda > 0$ is an adjustable (regularization) parameter.

Solution: From the definition, $\mathrm{prox}_{f_2}(\mathbf{x}) = \arg\min_{\mathbf{z}\in\mathbb{R}^d} J(\mathbf{z},\mathbf{x})$ where $J(\mathbf{z},\mathbf{x})$ may be rewritten as

$$J(\mathbf{z},\mathbf{x}) = \tfrac{1}{2}(\mathbf{z}-\mathbf{x})^\mathsf{T}(\mathbf{z}-\mathbf{x}) + \tfrac{1}{2}\lambda\mathbf{z}^\mathsf{T}\mathbf{L}^\mathsf{T}\mathbf{L}\mathbf{z}.$$

The first observation is that $\mathbf{z} \mapsto J(\mathbf{z},\mathbf{x})$ is coercive (simply because the first term is coercive) and continuous, strictly-convex because it is the sum of two continuous and strictly-convex functionals. This allows us to deduce that the minimizer of (4) exists and is unique, in conformity with the theory of proximal operators. Next, we use the rules of differential vector calculus to determine the gradient of $J(\mathbf{z}, \mathbf{x})$ as

$$\frac{\partial J(\mathbf{z}, \mathbf{x})}{\partial \mathbf{z}} = (\mathbf{z} - \mathbf{x}) + \lambda \mathbf{L}^\top \mathbf{L} \mathbf{z}. \tag{6}$$

The minimizer is found by setting $\frac{\partial J(\mathbf{z}, \mathbf{x})}{\partial \mathbf{z}} = \mathbf{0} \;\Leftrightarrow\; \mathbf{x} = (\mathbf{I} + \lambda \mathbf{L}^\top \mathbf{L})\mathbf{z}$, which then yields

$$\mathbf{z} = \mathrm{prox}_f(\mathbf{x}) = (\mathbf{I} + \lambda \mathbf{L}^\top \mathbf{L})^{-1} \mathbf{x}. \tag{7}$$

We observe that the matrix $\mathbf{A} = \mathbf{I} + \lambda \mathbf{L}^\top \mathbf{L}$ is symmetric positive-definite, and invertible: its eigenvalues are all greater then 1 because of the leading identity matrix. This is one more confirmation that the proximal operator is well defined and single-valued.

It is instructive to compare this operator with the variational Tikhonov denoiser investigated in Exercise 7.1b. Specifically, if $\mathbf{L}$ is a convolution matrix, then prox_{f_2} is the discrete equivalent of the Tikhonov/Wiener denoiser $g \mapsto r_{\mathrm{Tik}} * g$ and (7) has a fast matrix-free implementation that involves the FFT.