Chapter 10

Signal models and basis functions

ch,basis

Contents

10.1 Introduction (s,basis,intro) **10.1**
10.2 (Linear) basis function choices (s,basis,linear) **10.3**
10.3 Nonlinear object parameterizations (s,basis,nonlin) **10.3**
10.4 Synthesis models and dictionary learning (s,basis,synth) **10.5**
10.4.1 Subspace model 10.5
10.4.2 Union of subspaces models 10.6
10.4.3 Over-complete dictionary model 10.6
10.4.4 Dictionary learning 10.7
10.4.4.1 Dictionary learning: image domain 10.8
10.4.4.2 Image reconstruction using a dictionary 10.9
10.4.4.3 Dictionary learning: data domain 10.10
10.4.5 Summary 10.10
10.5 Bibliography **10.11**

s,basis,intro

10.1 Introduction (s,basis,intro)

Most of the reconstruction methods described in this *book* assume that the underlying continuous-space object of interest is approximated by a *series expansion* in terms of basis functions [1]:

e,basis,intro,series

$$f(\vec{x}) = \sum_{j=1}^{n_p} x_j \, b_j(\vec{x}), \tag{10.1.1}$$

or equivalently

$$f = \mathcal{B}_\square \boldsymbol{x},$$

where $\mathcal{B}_\square$ is a mapping from $\mathbb{R}^{n_p}$ or $\mathbb{C}^{n_p}$ into the space of functions of $\vec{x}$. The jth "column" of the operator $\mathcal{B}_\square$ is the basis function $b_j(\vec{x})$. Although we often write equality in the model (10.1.1), in practice this representation is an *approximation*. See Fig. 10.1.1.

The model (10.1.1) is equivalent to assuming that $f(\vec{x})$ lies in the n_p-dimensional *subspace* of continuous-space functions spanned by the basis functions $\{b_j(\vec{x})\}$. We usually use such finite-dimensional subspaces because it facilitates computation and because the available data vector $\boldsymbol{y}$ is *always* finite dimensional.

Nevertheless, in principle it is at least conceivable to try to reconstruct a continuous $f(\vec{x})$ from a finite-dimensional measurement vector. This is done routinely in the field of nonparametric regression [2], the generalization of linear regression that allows for fitting smooth functions discussed in 1D in §2.2. However, nonparametric estimation becomes more complicated in 2D problems like tomography.

Van De Walle, Barrett, *et al.* [3] proposed a pseudo-inverse approach to MRI reconstruction in a continuous-object / discrete-data formulation, based on the general principles of Bertero *et al.* [4]. If the pseudo-inverse could truly be computed once-and-for-all then such an approach could be practically appealing. However, in practice often there are object-dependent effects, such as nonuniform attenuation in SPECT and magnetic field inhomogeneity in MRI, and these effects preclude precomputation of the required SVDs. So pseudo-inverse approaches are impractical computationally for typical realistic physical models. See also [5, 6].

This chapter summarizes some of the choices for basis functions $\{b_j(\vec{x})\}$ that have been investigated in the image reconstruction literature, as well as some of the alternatives to the finite-dimensional linear subspace model (10.1.1).

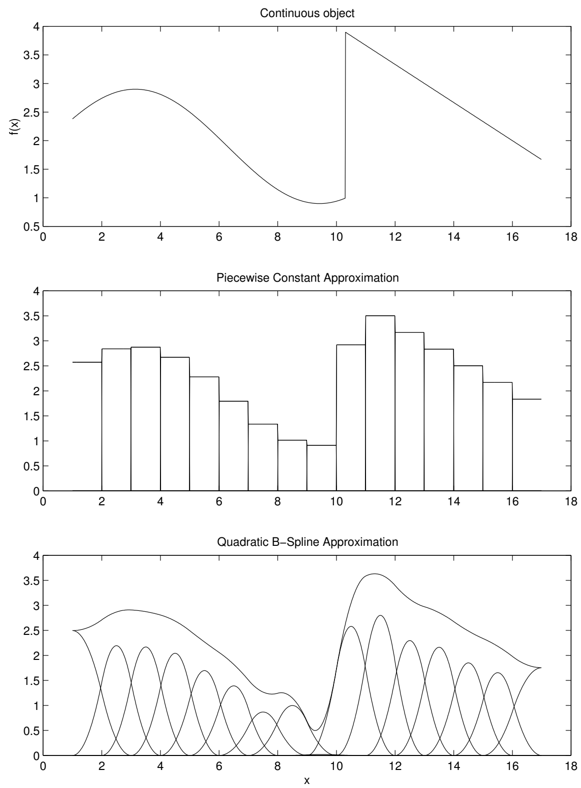


Figure 10.1.1: Illustration of a 1D function $f(x)$ and its approximation by two finite-dimensional subspaces, one using rectangular basis functions, and the other using quadratic B-splines. fig_basis1

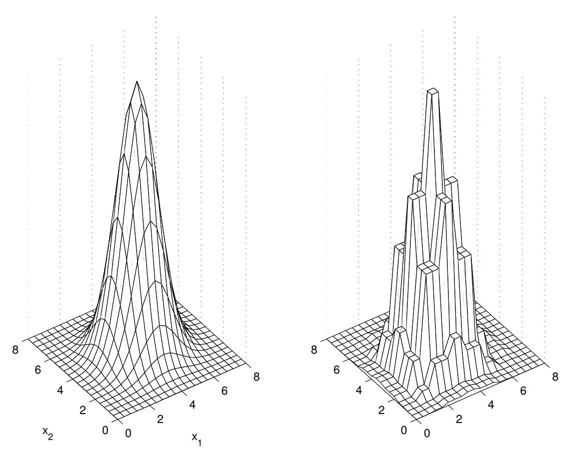


Figure 10.1.2: Illustration of a simple 2D function $f(\vec{x})$ and its approximation by a finite-dimensional subspaces using square-pixel basis functions. fig_basis2

s,basis,linear
10.2 (Linear) basis function choices (s,basis,linear)

Numerous families of basis functions $\{b_j(\vec{x})\}$ have been investigated in the image reconstruction literature for use with finite-dimensional linear subspace models of the form (10.1.1). This section enumerates several of the options.

- Fourier series (complex / not sparse) [7]
- Circular harmonics (complex / not sparse) [8–11]
- Wavelets (negative values / not sparse) [12]
- Overlapping circles (disks) [13]
- Overlapping spheres (balls) [14] (approximately ellipsoids under nonrigid motion)
- Kaiser-Bessel window functions (blobs) [15]
- Rectangular pixels / voxels (rect functions)
- Dirac impulses (point masses / bed-of-nails / lattice of points / "comb" function)
- "Natural pixels" $\{\mathsf{s}_i(\vec{x})\}$ [16–20]
- B-splines (pyramids) [21–24]
- Polar grid [25–29],
- Logarithmic polar grid [30]
- gaussian functions [31, 32]
- *Radial basis functions* (circularly symmetry) such as [33] $b(\vec{x}) = (1 - \|\vec{x}/r\|^2) 1_{\{\|\vec{x}\| \leq r\}}$
- Organ-based voxels (*e.g.*, for PET reconstruction using anatomy from CT image in a PET-CT system) [34–39]

There are many considerations when choosing between the many options listed above. (See [22, 40] for early discussions.)

Mathematical considerations

- The subspace should represent $f(\vec{x})$ "well" with moderate n_{p} (approximation accuracy). One meaning of "well" is that the *approximation error* should be much less than the *estimation error*.
- In particular, it is desirable for the subspace to be able to represent perfectly a constant (uniform) function.
- It is desirable for the functions $\{b_j(\vec{x})\}$ to be *linearly independent*, ensuring that the expansion (10.1.1) is unique. Uniqueness is not essential though because our ultimate goal is finding f^{true} and having multiple values of $\boldsymbol{x}$ for which $f^{\mathrm{true}} = \boldsymbol{\mathcal{B}}_{\square} \boldsymbol{x}$ (or approximately so) is acceptable. There is really never a $\boldsymbol{x}_{\mathrm{true}}$ in practice! However, the redundancy of linearly dependent functions $\{b_j(\vec{x})\}$ may increase computation time so usually should be avoided.
- Orthogonality of basis functions is often advocated for signal modeling because it greatly simplifies computing the coefficients x_j in (10.1.1) given f. However, in inverse problems we are not given f so orthogonality is not essential!
- It can be desirable for the representation (10.1.1) to be insensitive to a shift of the basis-function grid, *i.e.*, approximate *shift invariance*. Smooth basis functions are preferable to discontinuous functions in this respect.
- Similarly, it is desirable for the basis functions to be *symmetric* and *circularly symmetric* so that the representation (10.1.1) has approximate *rotation invariance* and invariance to coordinate system reflections.
- Many bases have the desirable approximation property that one can form arbitrarily accurate approximations to $f(\vec{x})$ by taking n_{p} sufficiently large. (This is related to *completeness*.) Exceptions include "natural pixels" (a finite set) and the point-lattice "basis" (usually).

Computational considerations

- Choice of basis functions affects how "easy" it is to compute elements of the system matrix a_{ij} and/or perform matrix vector multiplication $\boldsymbol{A}\boldsymbol{x}$ (forward projection).
- If the system matrix $\boldsymbol{A}$ is to be precomputed and stored, then it should be sparse (mostly zeros). Narrower basis functions usually are preferable in this respect.
- In applications where $f(\vec{x})$ is nonnegative, it should be easy to represent nonnegative functions *e.g.*, if $x_j \geq 0$, then $f(\vec{x}) \geq 0$. A sufficient condition is $b_j(\vec{x}) \geq 0$.

As detailed in Appendix 25, basis function choice, combined with the system model $\bar{y}_i = \int \mathsf{s}_i(\vec{x}) \, f(\vec{x}) \, \mathrm{d}\vec{x}$ determines the elements of the system matrix $\boldsymbol{A}$ as $a_{ij} = \int \mathsf{s}_i(\vec{x}) \, b_j(\vec{x}) \, \mathrm{d}\vec{x}$. However, many published "projector / backprojector pairs" are not based explicitly on any particular choice of basis.

Some pixel-driven backprojectors could be interpreted implicitly as point-mass object models. This model works fine for FBP, but causes artifacts for iterative methods [41].

Mazur *et al.* [42] approximate the shadow of each pixel by a rect function, instead of by a trapezoid. "As the shapes of pixels are artifacts of our digitisation of continuous real-world images, consideration of alternative orientation or shapes for them seems reasonable." However, they observe slightly worse results that worsen with iteration!

s,basis,nonlin
10.3 Nonlinear object parameterizations (s,basis,nonlin)

Linear models like (10.1.1) are the most common in image reconstruction, but numerous alternatives have also been investigated, including nonlinear parametric models. These models are often considered in problems with very limited data, necessitating strong assumptions about the object with few degrees of freedom. Usually these models involve estimating both object intensity *and* shape parameters (*e.g.*, location, radius, etc.), the latter leading to nonlinearity. This section enumerates some of the parametric models that have been used for image reconstruction problems.

Surface-based (homogeneous) models

- Circles / spheres [43,44]
- Ellipses / ellipsoids [45]
- Superquadrics
- Rectangles with unknown position, size, amplitude [46,47]
- Polygons [48]
- Bi-quadratic triangular Bezier patches [49]
- Triangulated 3D surface [50–52].

Other models

- Generalized series $f(\vec{x}) = \sum_j x_j \, b_j(\vec{x}, \boldsymbol{\theta})$
- Deformable templates [53–55] $f(\vec{x}) = b(T_{\boldsymbol{\theta}}(\vec{x}))$
- Adaptive tetrahedral meshes (defined by a point cloud) [56]

These type of parametric models have the advantage that they can be considerably more *parsimonious* than voxelized models. Thus, if they are accurate, they can yield greatly reduced *estimation error*. These models are particularly compelling in limited-data problems, even if they are oversimplified. (All models are wrong but some models are useful.) The nonlinear dependence of $f(\vec{x})$ on shape parameters and location leads to non-convex cost functions, complicating optimization.

Region of interest (ROI) or "focus of attention" [57–59] s,basis,synth

10.4 Synthesis models and dictionary learning (s,basis,synth)

Even after using the finite-series expansion (10.1.1), image reconstruction problems are usually under-determined (*i.e.*, $n_\mathrm{p} > n_\mathrm{d}$) or at least very badly ill-conditioned. Thus, attempting to estimate $\boldsymbol{x}$ solely by finding a minimizer of a data-fit term $\text{Ł}(\boldsymbol{x})$ is rarely adequate. Chapter 1 and Chapter 2 described using *regularization* or *prior models* to (indirectly) "constrain" the estimate $\hat{\boldsymbol{x}}$. This section describes synthesis-based approaches to modeling $\boldsymbol{x}$. The basic idea underlying all of these methods is that we must find some way to reduce the *degrees of freedom* of the estimator $\hat{\boldsymbol{x}}$.

10.4.1 Subspace model

The classical linear approach to reducing degrees of freedom is to assume that $\boldsymbol{x}$ lies in the subspace $\mathcal{S}$ spanned by some $n_\mathrm{p} \times r$ basis matrix $\boldsymbol{B}$ with $r \ll n_\mathrm{p}$ (having full column rank):

e,basis,synth,x=Bz

$$\boldsymbol{x} \approx \boldsymbol{B}\boldsymbol{z}, \tag{10.4.1}$$

for some unknown coefficient vector $\boldsymbol{z} \in \mathbb{R}^r$. (This is equivalent mathematically to using a different set of r basis functions in (10.1.1).) Then one estimates $\boldsymbol{z}$ from the data by minimizing the data-fit term:

$$\hat{\boldsymbol{x}} \triangleq \boldsymbol{B}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} \triangleq \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathbb{R}^r} \text{Ł}(\boldsymbol{B}\boldsymbol{z}).$$

If r is sufficiently small relative to the data dimension n_d, then this problem may be well conditioned and have a unique minimizer.

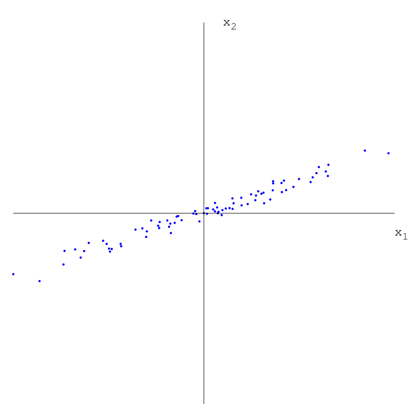
One can "learn" the subspace basis $\boldsymbol{B}$ from training data $\boldsymbol{X} = [\boldsymbol{x}_1 \ \dots \ \boldsymbol{x}_M]$, provided $M > r$. We want to find a matrix $\boldsymbol{B}$ (with linearly independent columns) such that each $\boldsymbol{x}_m$ approximately lies in its span, *i.e.*, $\|\boldsymbol{x}_m - \boldsymbol{B}\boldsymbol{z}_m\|$ should be small, where $z_m = \arg\min_{\boldsymbol{z}} \|\boldsymbol{x}_m - \boldsymbol{B}\boldsymbol{z}\|$. It is simpler (and loses no generality) to require that the columns of $\boldsymbol{B}$ be orthonormal, leading to the optimization problem

e,basis,synth,B,train

$$\operatorname*{arg\,min}_{\boldsymbol{B}\,:\,\boldsymbol{B}'\boldsymbol{B}=\boldsymbol{I}_r} \ \min_{\boldsymbol{Z}\in\mathbb{R}^{r\times M}} \|\boldsymbol{X} - \boldsymbol{B}\boldsymbol{Z}\|^2_{\mathrm{Frob}}. \tag{10.4.2}$$

This is a non-convex problem and there is not a unique solution for $\boldsymbol{B}$, but one of the global minimizers is to let $\boldsymbol{B}$ be the first r eigenvectors of the $n_\mathrm{p} \times n_\mathrm{p}$ sample covariance matrix $\boldsymbol{X}\boldsymbol{X}' = \sum_{m=1}^{M} \boldsymbol{x}_m \boldsymbol{x}_m'$, assuming the eigenvalues λ_k are ordered from largest to smallest. If $\lambda_r > \lambda_{r+1}$ then the subspace spanned by this $\boldsymbol{B}$ is unique even if $\boldsymbol{B}$ itself is not unique (one can permute its columns for example). But in the rare case where $\lambda_r = \lambda_{r+1}$, then the subspace is not unique but still $\boldsymbol{B}$ is a global minimizer. This approach is called *principal components analysis* (*PCA*). In practice, one computes $\boldsymbol{B}$ by finding the *singular value decomposition* (*SVD*) of $\boldsymbol{X}$ and taking the first r left singular vectors.

It is difficult to visualize subspaces in dimensions higher than $n_\mathrm{p} = 3$. Fig. 10.4.1 illustrates an $n_\mathrm{p} = 2$ case where a $r = 1$ dimensional subspace is a reasonable model.



fig_basis_synth_subspace1

Figure 10.4.1: Illustration of (synthetic) data that is well approximated by a 1D subspace.

The subspace approach works fine for regression problems like fitting a line ($r = 2$) to a scatter plot of data. Unfortunately, in image reconstruction problems rarely is there a single low-dimensional subspace that adequately describes the images of interest. The remainder of this section describes models that are more flexible yet retain some aspects of the "low rank" nature of the classical subspace model.

10.4.2 Union of subspaces models

A more general model is to assume $\boldsymbol{x}$ belongs (at least approximately) to one of K subspaces where the kth subspace is spanned by the columns of a $n_\mathrm{p} \times r_k$ matrix $\boldsymbol{B}_k$, where the rank $r_k \ll n_\mathrm{p}$, as follows:

$$\boldsymbol{x} \in \bigcup_{k=1}^{K} \mathcal{S}_k, \quad \mathcal{S}_k = \{\boldsymbol{B}_k \boldsymbol{z}_k : \boldsymbol{z}_k \in \mathbb{R}^{r_k}\}.$$

If we can somehow determine which subspace is appropriate, then essentially we have only a r_k-dimensional estimation problem within that subspace. One possible formulation is

$$\hat{\boldsymbol{x}} \triangleq \boldsymbol{B}_{\hat{k}} \hat{\boldsymbol{z}}_{\hat{k}}, \quad \hat{\boldsymbol{z}}_k \triangleq \arg\min_{\boldsymbol{z} \in \mathbb{R}^{r_k}} \mathsf{Ł}(\boldsymbol{B}_k \boldsymbol{z}), \quad \hat{k} \triangleq \arg\min_k \mathsf{Ł}(\boldsymbol{B}_k \hat{\boldsymbol{z}}_k).$$

Methods for learning the subspace bases $\{\boldsymbol{B}_k\}$ from training data are given in [60–63]. Often the models include hierarchical or tree-based structure [64].

A union of K-subspaces is a generalization of the K-means approach to data clustering.

Fig. 10.4.2 illustrates an $n_\mathrm{p} = 3$ case with $K = 3$ subspaces where $r_1 = 2$ and $r_2 = r_3 = 1$.

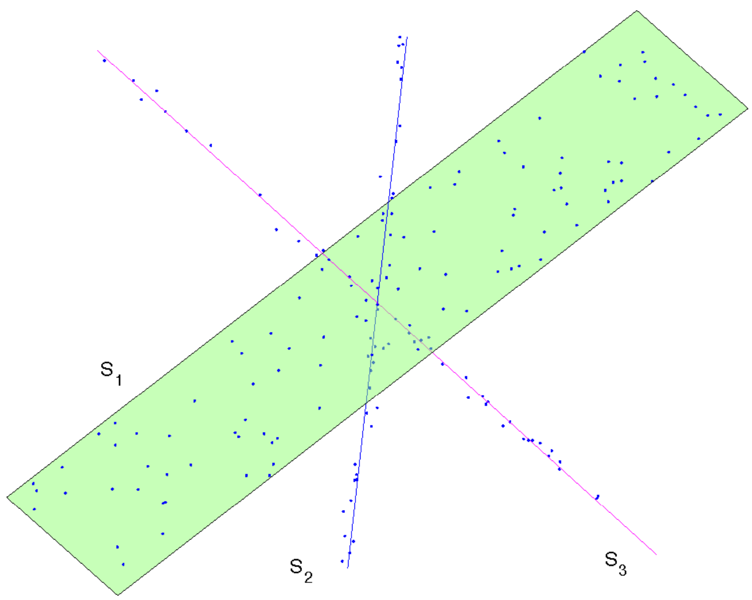


Figure 10.4.2: Illustration of (synthetic) data in $\mathbb{R}^{n_\mathrm{p}}$ where $n_\mathrm{p} = 3$ that is well approximated by $K = 3$ subspaces, the first of dimension $r_1 = 2$ and $r_2 = r_3 = 1$ for the other two subspaces. fig_basis_synth_subspace3

10.4.3 Over-complete dictionary model

An alternative to the union-of-subspaces model is to represent $\boldsymbol{x}$ using a linear combination of a small number of atoms from an *over-complete dictionary* $\boldsymbol{D}$ of size $n_\mathrm{p} \times K$ where $K > n_\mathrm{p}$ as follows [65]:

e,basis,synth,x=Dz

$$\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{z}, \quad \|\boldsymbol{z}\|_0 = \sum_{k=1}^{K} 1_{\{z_k \neq 0\}} \leq r. \tag{10.4.3}$$

In other words, we represent (or approximate) $\boldsymbol{x}$ using a linear model with an r-sparse coefficient vector. Certainly this model can be more expressive than the single subspace model (10.4.1). Strictly speaking, (10.4.3) is also a union of subspaces model where there are $\binom{K}{r}$ possible r-dimensional subspaces. This representation is “more flexible” yet “less structured.”

For a given dictionary $\boldsymbol{D}$, one can approach an image reconstruction problem as follows:

e,basis,synth,over,recon

$$\hat{\boldsymbol{x}} \triangleq \boldsymbol{D}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} \triangleq \arg\min_{\boldsymbol{z} \in \mathbb{R}^K} \mathsf{Ł}(\boldsymbol{D}\boldsymbol{z}) \text{ sub. to } \|\boldsymbol{z}\|_0 \leq r. \tag{10.4.4}$$

The idea here is that after we somehow determine the appropriate subspace (*i.e.*, the span of the columns of $\boldsymbol{D}$ corresponding to the non-zero elements of $\boldsymbol{z}$), then the minimization problem is “only” r-dimensional, where typically

$r \ll n_{\text{p}}$, so the problem may be well conditioned within that subspace. Note that the length of the coefficient vector $\boldsymbol{z}$ in the subspace model (10.4.1) is only r whereas in the over-complete dictionary model (10.4.3) it is $K \geq n_{\text{p}} \gg r$.

Fig. 10.4.3 illustrates an $n_{\text{p}} = 2$ case with $K = 3$ atoms where $r = 1$ provides a reasonable model. In the case $n_{\text{p}} = 2$, the union-of-subspaces model and the over-complete dictionary model are identical for $r = 1$. To illustrate the distinction between these models requires $n_{\text{p}} > 2$. For example, for $n_{\text{p}} = 3$, if we have a dictionary with K atoms and $r = 2$, then (10.4.3) corresponds to the union of all $\binom{K}{2}$ pairs of atoms in the dictionary. In contrast a typical union-of-subspaces model would consist of just a few such pairs.

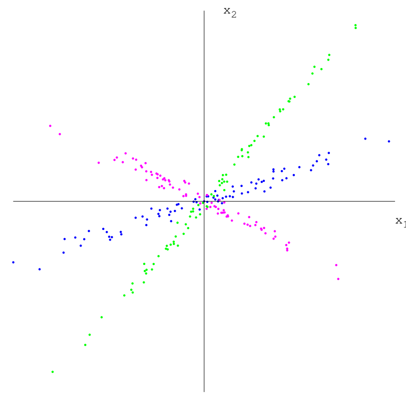


Figure 10.4.3: Illustration of (synthetic) data for $n_{\text{p}} = 2$ that is well approximated by $K = 3$ subspaces of dimension $r = 1$. fig_basis_synth_subspace2

There are many issues to address when using an over-complete dictionary like (10.4.3).

- One must choose the allowed sparsity level r, which is rarely known *a priori* for a given application.
  For problems with gaussian noise of known variance σ^2, one can avoid picking r by replacing the optimization problem (10.4.4) with the following alternative formulation:

$$\hat{\boldsymbol{x}} \triangleq \boldsymbol{D}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} \triangleq \underset{\boldsymbol{z} \in \mathbb{R}^K}{\arg\min} \|\boldsymbol{z}\|_0 \text{ sub. to } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2^2 \leq n_{\text{d}}\sigma^2.$$

  Here we find the sparsest coefficient vector that leads to a data-fit term within a given error tolerance. The tolerance $n_{\text{d}}\sigma^2$ is appropriate if $\boldsymbol{x}_{\text{true}} = \boldsymbol{D}\boldsymbol{z}$, but we have seen in §2.5.2.1 that in general the *discrepancy principle* yields suboptimal regularization levels when fitting data.
- The minimization problem in (10.4.3) is non-convex and combinatorial. Often this problem is circumvented by using greedy algorithms or replacing $\|\boldsymbol{z}\|_0$ with $\|\boldsymbol{z}\|_1$ leading to a convex optimization problem.
- The size of the dictionary $\boldsymbol{D}$ would be enormous if it were used to represent the entire image. For a 256^2 image (tiny by modern digital photography standards), $n_{\text{p}} = 2^{16}$ and if $K = 2n_{\text{p}}$ then $\boldsymbol{D}$ has 2^{33} elements and storing $\boldsymbol{D}$ as single-precision (4-byte) floating point values would need 2^{35} bytes or 32 GBytes. Therefore usually the image $\boldsymbol{x}$ is partitioned into small patches, *e.g.*, each of 8×8 pixels, and each patch of 64 values is represented individually by a sparse coefficient vector for an over-complete dictionary $\boldsymbol{D}_0$ of size $8^2 \times K$ where $K > 8^2$. If $K = 128$ then only 32 KBytes are needed to store $\boldsymbol{D}_0$. Mathematically, in this representation the overall model for $\boldsymbol{x}$ corresponds to the *direct sum* [wiki] of the subspaces for each image patch.
  Issues with patch-wise representations include possible image artifacts at the boundaries between patches, leading to "blocky" appearance, particularly if r is small. Furthermore, this representation is not *shift invariant*.
  To avoid these drawbacks, one could use overlapping patches, and then one must choose how to combine the values from each patch to make the final image estimate $\hat{\boldsymbol{x}}$.
- Finally, one must select the dictionary $\boldsymbol{D}$, including K, the number of atoms. The next section describes a popular method called the *K-SVD* approach.

10.4.4 Dictionary learning

In the context of image reconstruction, there are two main types of methods for *dictionary learning*, *i.e.*, for estimating $\boldsymbol{D}$. One approach is to start with image-domain training data, typically obtained from "fully sampled" and/or "high SNR" cases, and then we use the learned dictionary $\boldsymbol{D}$ to help reconstructed "under sampled" and/or "low SNR" data. The other approach is to try to learn the dictionary $\boldsymbol{D}$ adaptively from the given measurement vector $\boldsymbol{y}$ (sinogram, k-space, etc.) [66, 67]. The following sections consider both types of methods, starting with the simpler the image-domain approach.

10.4.4.1 Dictionary learning: image domain

Given training data $\boldsymbol{X} = [\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_M]$ (*e.g.*, patches from good quality images), we can *learn* a dictionary $\boldsymbol{D}$ using the *K-SVD* method [68]. We start by picking K; clearly we need $M \gg K$ or the dictionary learning problem would degenerate to just using the training data.

The goal is to find a dictionary $\boldsymbol{D} = [\boldsymbol{d}_1 \ \ldots \ \boldsymbol{d}_K]$ such that every example in the training data is well approximated by the dictionary with a r-sparse coefficient vector. In other words, for each $\boldsymbol{x}_m$ there should be some coefficient vector $\boldsymbol{z}_m$ with $\|\boldsymbol{z}_m\|_0 \leq r$ for which $\|\boldsymbol{x}_m - \boldsymbol{D}\boldsymbol{z}_m\|$ is small. A reasonable mathematical criterion is

$$\mathop{\arg\min}_{\boldsymbol{D} \in \mathbb{R}^{n_\mathrm{p} \times K}} \sum_{m=1}^{M} \min_{\boldsymbol{z}_m \in \mathcal{Z}_r} \|\boldsymbol{x}_m - \boldsymbol{D}\boldsymbol{z}_m\|_2^2, \quad \mathcal{Z}_r \triangleq \left\{ \boldsymbol{z} \in \mathbb{R}_m^K : \|\boldsymbol{z}\|_0 \leq r \right\}.$$

Often this minimization problem is written concisely by grouping together the coefficient vectors $\boldsymbol{Z} \triangleq [\boldsymbol{z}_1 \ \ldots \ \boldsymbol{z}_M]$ and writing

$$\mathop{\arg\min}_{\boldsymbol{D} \in \mathbb{R}^{n_\mathrm{p} \times K}} \min_{\boldsymbol{Z}} \|\!|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{Z}|\!\|_{\mathrm{Frob}}^2 \text{ sub. to } \boldsymbol{z}_m \in \mathcal{Z}_r \ \forall m.$$

This expression is slightly more concise, but somewhat obscures the fact that for a given $\boldsymbol{D}$, we can compute each coefficient vector $\boldsymbol{z}_m$ in parallel (independently).

The K-SVD approach [68] alternates between updating the coefficients $\boldsymbol{Z}$ for a given candidate dictionary $\boldsymbol{D}$, and then updating *sequentially* each atom $\boldsymbol{d}_m$ in the dictionary $\boldsymbol{D}$ using an SVD. This *alternating minimization* method decreases the cost function $\|\!|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{Z}|\!\|_{\mathrm{Frob}}$ every iteration and thus the cost function values converge to some non-negative value. However, contrary to the claims in [68], this monotonicity is insufficient to ensure convergence of $\boldsymbol{D}$ to a local minimizer without further proof. (See Example 11.1.2.)

For a given candidate dictionary $\boldsymbol{D}$ in this alternating minimization process, the problem of updating $\boldsymbol{Z}$ has many names in the literature, including *sparse coding*, *sparse approximation*, *sparse synthesis*, and *atom decomposition*. All of these terms refer to the following minimization problem:

e,basis,synth,code

$$\hat{\boldsymbol{z}}_m \triangleq \mathop{\arg\min}_{\boldsymbol{z} \in \mathcal{Z}_r} \|\boldsymbol{x}_m - \boldsymbol{D}\boldsymbol{z}\|_2^2, \quad m = 1, \ldots, M. \tag{10.4.5}$$

Methods for solving this combinatorial optimization problem are called *pursuit algorithms*. The simplest methods are *greedy algorithms* that select one atom at a time. See [69] for a survey.

The *matching pursuit* (*MP*) method chooses the atom having the largest (absolute) inner product with the residual, as follows [70] [wiki].

Matching pursuit

Input: Dictionary $\boldsymbol{D} = [\boldsymbol{d}_1 \ \ldots \ \boldsymbol{d}_K]$ and signal vector $\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{p}}$.
$\boldsymbol{r} = \boldsymbol{x}$ (initialize residual)
$\boldsymbol{z} = \boldsymbol{0}_K$ (initialize sparse coefficient vector)
for $n = 1, \ldots, r$

$$\begin{aligned} k_n &= \mathop{\arg\max}_{k} |\langle \boldsymbol{d}_k, \boldsymbol{r} \rangle| && \text{(atom index selection)} \\ z_{k_n} &:= \langle \boldsymbol{d}_{k_n}, \boldsymbol{r} \rangle && \text{(coefficient)} \\ \boldsymbol{r} &:= \boldsymbol{r} - z_{k_n} \boldsymbol{d}_{k_n} && \text{(updated residual)} \end{aligned}$$

end
Output: atom indices $k_1, \ldots, k_r$ and coefficients $z_{k_1}, \ldots, z_{k_r}$ such that $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{z} = \sum_{n=1}^{r} z_{k_n} \boldsymbol{d}_{k_n}$

The simple inner product $\langle \boldsymbol{d}_{m_k}, \boldsymbol{r} \rangle$ provides an appropriate coefficient if the selected atoms are orthonormal, but if they are not then one can obtain a better fit (smaller residual) more generally performing a least-squares fit of the coefficients after each new atom is selected. Equivalently, we project the residual onto the span of all selected atoms at each step. The *orthogonal matching pursuit* (*OMP*) method uses this variation, along with the (natural) constraint that each atom can be picked only once [71] [wiki].

Being greedy algorithms, they typically find local minimizers of the sparse coding problem (10.4.5). An alternative is to replace the non-convex problem (10.4.5) with one of the following convex problems:

$$\min_{\boldsymbol{z} \in \mathbb{R}^K} \|\boldsymbol{z}\|_1 \text{ sub. to } \|\boldsymbol{x}_m - \boldsymbol{D}\boldsymbol{z}\|_2^2 \leq \varepsilon$$

e,basis,synth,bpd

$$\min_{\boldsymbol{z} \in \mathbb{R}^K} \frac{1}{2} \|\boldsymbol{x}_m - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \beta \|\boldsymbol{z}\|_1. \tag{10.4.6}$$

These optimization problems are called *basis pursuit denoising* [wiki]. As $\beta \to 0$ the problems become *basis pursuit* [wiki]. To use (10.4.6) for updating $\boldsymbol{Z}$ for the K-SVD method, one can take the r largest non-zero coefficients of $\hat{\boldsymbol{z}}$, or adjust β so that $\hat{\boldsymbol{z}}$ has r non-zero coefficients.

Having updated the sparse coefficients $\boldsymbol{Z}$, the next step of the K-SVD method is to update the dictionary $\boldsymbol{D}$. The K-SVD uses block coordinate descent to update $\boldsymbol{D}$ by minimizing $\|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{Z}\|_{\text{Frob}}$ with respect to one column of $\boldsymbol{D}$ (and the corresponding non-zero elements of $\boldsymbol{Z}$) sequentially. To update $\boldsymbol{d}_k$, we consider all training samples for which the kth element of the corresponding coefficient vector is nonzero, *i.e.*, $\mathcal{M}_k \triangleq \{m : z_{mk} \neq 0\}$. We then compute the residual associated with those training samples, excluding $\boldsymbol{d}_k$, *i.e.*,

$$\boldsymbol{r}_m \triangleq \sum_{j \neq k} z_{mj} \boldsymbol{d}_j, \quad m \in \mathcal{M}_k.$$

We then find a new vector $\boldsymbol{d}_k$ that best spans these residual vectors by taking the first singular vector of the SVD of $\{\boldsymbol{r}_m\}$. This step is analogous to solving (10.4.2) with $r = 1$. The corresponding elements of $\boldsymbol{Z}$ are then updated with this new atom $\boldsymbol{d}_k$, without changing the sparsity pattern of $\boldsymbol{Z}$. This process is repeated for $k = 1, \ldots, K$, after which the algorithm returns to the sparse coding step to update $\boldsymbol{Z}$. Several implementation tricks related to initialization and pruning are given in [68], such as always including the constant vector $\mathbf{1}_{n_\text{p}}$ as one of the atoms.

Fig. 10.4.4, taken from [68, Fig. 5], compares a dictionary learned from 500 8×8 patches from face images, compared to Harr and DCT bases.

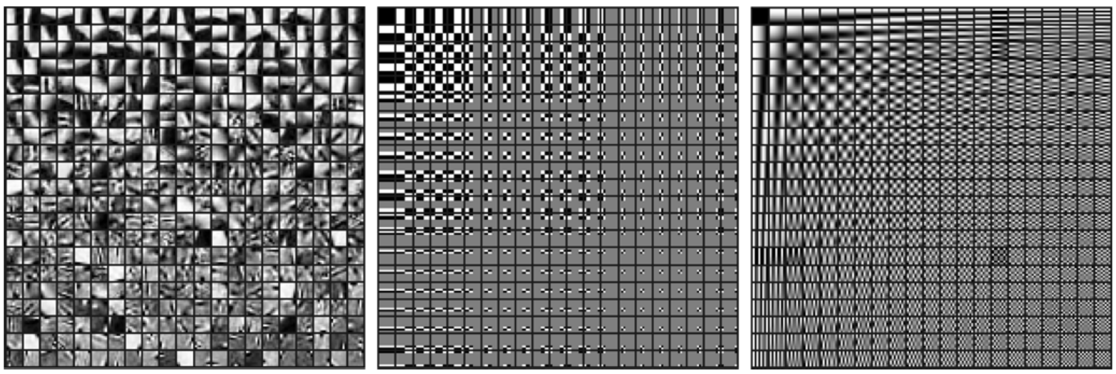


Figure 10.4.4: A dictionary for 8×8 patches learned from face images, compared to (over-complete) Harr and DCT bases. aharon-06-ksa,fig5

For large M (many training samples) one can use stochastic gradient descent [72]. For some applications, constraints such as nonnegativity on the atoms is appropriate [73].

For improvements see [74–76].

For software, see http://www.cs.technion.ac.il/~elad/software.

10.4.4.2 Image reconstruction using a dictionary

Having selected a dictionary $\boldsymbol{D}$, *e.g.*, by learning it from training data, the image reconstruction problem is to estimate $\boldsymbol{x}$ from data $\boldsymbol{y}$ using the sparsity model (10.4.3). There are several possible approaches.

The traditional synthesis approach is to estimate the coefficients as follows:

$$\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_p \text{ sub. to } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2^2 \leq \varepsilon.$$

Instead of requiring that $\boldsymbol{x}$ be synthesized exactly from the dictionary, another option is to encourage $\hat{\boldsymbol{x}}$ to be similar to such an image: [67]:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}) = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}^{1/2}}^2 + \beta \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2 \text{ sub. to } \|\boldsymbol{z}\|_0 \leq r.$$

Another option is the following convex relaxation of the previous problem [67]:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}) = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}^{1/2}}^2 + \beta \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \alpha \|\boldsymbol{z}\|_1.$$

For these last two formulations, a natural optimization approach is to alternate between updating the image $\boldsymbol{x}$, which is a quadratic problem, and the coefficient vector $\boldsymbol{z}$, which is a sparse coding problem. (One could apply OMP or FISTA.) For further examples, see [77–83]. Interestingly, the convex form above is closely related to the augmented Lagrangian for the convex (but non-smooth) optimization problem

$$\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_{\boldsymbol{W}^{1/2}}^2 + \alpha \|\boldsymbol{z}\|_1.$$

10.4.4.3 Dictionary learning: data domain

A potential drawback of any dictionary learned from training data is that the object $\boldsymbol{x}$ being reconstructed might not have a sparse representation in that dictionary. Furthermore, there will always be noise in training data that will affect dictionary learning. Such considerations have motivated research on jointly reconstructing an image and learning a dictionary (or part of a dictionary) [66, 67, 77]. An example formulation is

$$\arg\min_{\boldsymbol{x}} \min_{\boldsymbol{D},\boldsymbol{Z}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \sum_{l=1}^{L} \beta \frac{1}{2} \|\boldsymbol{R}_l \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_l\|_2^2 \text{ sub. to } \|\boldsymbol{z}_l\|_0 \le r,$$

where $\boldsymbol{R}_l$ is a matrix that extracts the lth patch from the image $\boldsymbol{x}$. In this setting it is *essential* to use the dictionary to represent patches rather than the entire image. One can imagine many variations (including convex relaxations) of such formulations, as well as combining adaptively learned dictionaries with dictionaries learned from training data and some predetermined atoms (such as a DC component); one could also include conventional analysis regularizers, perhaps with a small regularization parameter.

An formulation that is convex with respect to each of the parameters individually (but not collectively) is [84]:

$$\arg\min_{\boldsymbol{x}} \min_{\boldsymbol{D},\boldsymbol{Z}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \sum_{l=1}^{L} \left(\beta \frac{1}{2} \|\boldsymbol{R}_l \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_l\|_2^2 + \alpha \|\boldsymbol{z}_l\|_1 \right).$$

To avoid scale ambiguity, here one must constrain $\boldsymbol{D}$, *e.g.*, by requiring that each column $\boldsymbol{d}_k$ have unit norm. An alternating minimization approach is natural.

One can also learn transforms for analysis formulations [85–87].

10.4.5 Summary

Unions of subspaces, such as over-complete dictionaries with sparse coefficient vectors, provide a mechanism for dimension reduction that is somewhat akin to classical subspaces but more flexible. The drawback is that the resulting optimization problems often have non-convex aspects.

Analysis formulations are "negative" in the sense that they *discourage* the reconstructed image $\hat{\boldsymbol{x}}$ from departing from prior assumptions about the relationships between groups of pixels (such as neighbors). In contrast, synthesis formulations are "positive" in the sense that they *encourage* (or require) the reconstructed image $\hat{\boldsymbol{x}}$ to be expressible in terms of the prior information (*e.g.*, linear combinations of dictionary atoms). Identifying which formulation (or combination thereof) is best for a given application is an *open problem*; see [88, 89].

10.5 Bibliography

censor:83:fse [1] Y. Censor. Finite series expansion reconstruction methods. *Proc. IEEE*, 71(3):409–19, March 1983. 10.1

silverman:85:sao [2] B. W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Royal Stat. Soc. Ser. B*, 47(1):1–52, 1985. 10.1

vandewalle:00:rom [3] R. Van de Walle, H. H. Barrett, K. J. Myers, M. I. Altbach, B. Desplanques, A. F. Gmitro, J. Cornelis, and I. Lemahieu. Reconstruction of MR images from data acquired on a general non-regular grid by pseudoinverse calculation. *IEEE Trans. Med. Imag.*, 19(12):1160–7, December 2000. 10.1

bertero:85:lip [4] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data, I: General formulation and singular system analysis. *Inverse Prob.*, 1(4):301–30, November 1985. 10.1

fall:11:cst [5] M. D. Fall, E. Barat, C. Comtat, T. Dautremer, T. Montagu, and S. Stute. Continuous space-time reconstruction in 4D PET. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2581–6, 2011. 10.1

fall:11:adc [6] M. D. Fall, E. Barat, C. Comtat, T. Dautremer, T. Montagu, and A. Mohammad-Djafari. A discrete-continuous Bayesian model for emission tomography. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 1373–6, 2011. 10.1

barrett:94:cbt [7] H. H. Barrett and H. C. Gifford. Cone-beam tomography with discrete data sets. *Phys. Med. Biol.*, 39(3):451–76, March 1994. 10.3

cormack:63:roa [8] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *J. Appl. Phys.*, 34(9):2722–7, September 1963. 10.3

cormack:64:roa [9] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications II. *japplphy*, 35(10):2908–13, October 1964. 10.3

hansen:81:toc [10] E. W. Hansen. Theory of circular harmonic image reconstruction. *J. Opt. Soc. Am.*, 71(3):304–8, March 1981. 10.3

sahiner:95:roi [11] B. Sahiner and A. E. Yagle. Region-of-interest tomography using exponential radial sampling. *IEEE Trans. Im. Proc.*, 4(8):1120–7, August 1995. 10.3

kisilev:01:wra [12] P. Kisilev, M. Zibulevsky, and Y. Zeevi. Wavelet representation and total variation regularization in emission tomography. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 1, pages 702–5, 2001. 10.3

shepp:82:mlr [13] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–22, October 1982. 10.3

reyes:07:mbr [14] M. Reyes, G. Malandain, P. M. Koulibaly, M. A. González-Ballester, and J. Darcourt. Model-based respiratory motion compensation for emission tomography image reconstruction. *Phys. Med. Biol.*, 52(12):3579–600, June 2007. 10.3

lewitt:90:mdi [15] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *J. Opt. Soc. Am. A*, 7(10):1834–46, October 1990. 10.3

minerbo:79:mam [16] G. Minerbo. MENT: A Maximum entropy algorithm for reconstructing a source from projection data. *Comp. Graphics and Im. Proc.*, 10(1):48–68, May 1979. 10.3

durrani:80:otf [17] T. S. Durrani and C. E. Goutis. Optimisation techniques for digital image recosntruction from their projections. *Proc. IEE–Comput. Digital Tech., pt. E*, 10(1):48–68, 1980. 10.3

natterer:80:eio [18] F. Natterer. Efficient implementation of optimal algorithms in computerized tomography. *Mathematical Methods in the Applied Sciences*, 2:545–55, 1980. 10.3

buonocore:81:anp [19] M. H. Buonocore, W. R. Brody, and A. Macovski. A natural pixel decomposition for two-dimensional image reconstruction. *IEEE Trans. Biomed. Engin.*, 28(2):69–78, February 1981. 10.3

hsieh:98:psi [20] Y-L. Hsieh, G. L. Zeng, and G. T. Gullberg. Projection space image reconstruction using strip functions to calculate pixels more ”natural” for modeling the geometric response of the SPECT collimator. *IEEE Trans. Med. Imag.*, 17(1):24–44, February 1998.

delaney:96:afa [21] A. H. Delaney and Y. Bresler. A fast and accurate Fourier algorithm for iterative parallel-beam tomography. *IEEE Trans. Im. Proc.*, 5(5):740–53, May 1996. 10.3

hanson:85:lbf [22] K. M. Hanson and G. W. Wecksung. Local basis-function approach to computed tomography. *Appl. Optics*, 24(23):4028–39, December 1985. 10.3

horbelt:02:dot [23] S. Horbelt, M. Liebling, and M. Unser. Discretization of the Radon transform and of its inverse by spline convolutions. *IEEE Trans. Med. Imag.*, 21(4):363–76, April 2002. 10.3

momey:11:anr [24] F. Momey, L. Denis, C. Mennessier, E. Thiebaut, J. M. Becker, and L. Desbat. A new representation and projection model for tomography, based on separable B-splines. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2602–9, 2011.

hebert:88:fmf [25] T. Hebert, R. Leahy, and M. Singh. Fast MLE for SPECT using an intermediate polar representation and a stopping criterion. *IEEE Trans. Nuc. Sci.*, 35(1):615–9, February 1988. 10.3

kaufman:87:iaa [26] L. Kaufman. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 6(1):37–51, March 1987. 10.3

israeljost:06:psi [27] V. Israel-Jost, P. Choquet, S. Salmon, C. Blondet, E. Sonnendrucker, and A. Constantinesco. Pinhole SPECT imaging: compact projection/backprojection operator for efficient algebraic reconstruction. *IEEE Trans. Med. Imag.*, 25(2):158–67, February 2006. 10.3

goussardy:13:ccr
[28] Y. Goussardy, A. Wagner, and M. Golkar. Cylindrical coordinate representation for statistical 3D CT reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 138–41, 2013. 10.3

thibaudeau:13:f3i
[29] C. Thibaudeau, J-D. Leroux, Réjean Fontaine, and R. Lecomte. Fully 3D iterative CT reconstruction using polar coordinates. *Med. Phys.*, 40(11):111904, November 2013.

eggermont:83:tro
[30] P. B. Eggermont. Tomographic reconstruction on a logarithmic polar grid. *IEEE Trans. Med. Imag.*, 2(1):40–8, March 1983. correction sep 1983. 10.3

kiebel:00:aib
[31] S. J. Kiebel, R. Goebel, and K. J. Friston. Anatomically informed basis functions. *NeuroImage*, 11(6):656–67, June 2000. 10.3

hamill:02:irm
[32] J. Hamill and T. Bruckbauer. Iterative reconstruction methods for high-throughput PET tomographs. *Phys. Med. Biol.*, 47(15):2627–36, August 2002. 10.3

perry:01:chs
[33] J. L. Perry and T. D. Gamble. Continuous high speed tomographic imaging system and method, 2001. US Patent 6,236,709. 10.3

snyder:84:usi
[34] D. L. Snyder. Utilizing side information in emission tomography. *IEEE Trans. Nuc. Sci.*, 31(1):533–7, February 1984. 10.3

carson:85:aml
[35] R. E. Carson, M. V. Green, and S. M. Larson. A maximum likelihood method for calculation of tomographic region-of-interest (ROI) values. *J. Nuc. Med. (Abs. Book)*, 26:P20, 1985. 10.3

carson:85:tep
[36] R. E. Carson and K. Lange. The EM parametric image reconstruction algorithm. *J. Am. Stat. Assoc.*, 80(389):20–2, March 1985. 10.3

carson:86:aml
[37] R. E. Carson. A maximum likelihood method for region-of-interest evaluation in emission tomography. *J. Comp. Assisted Tomo.*, 10(4):654–63, July 1986. 10.3

formiconi:93:lsa
[38] A. R. Formiconi. Least squares algorithm for region-of-interest evaluation in emission tomography. *IEEE Trans. Med. Imag.*, 12(1):90–100, March 1993. 10.3

reutter:98:kpe
[39] B. W. Reutter, G. T. Gullberg, and R. H. Huesman. Kinetic parameter estimation from attenuated SPECT projection measurements. *IEEE Trans. Nuc. Sci.*, 45(6):3007–13, December 1998.

herman:76:ira
[40] G. T. Herman and A. Lent. Iterative reconstruction algorithms. *Computers in Biology and Medicine*, 6(4):273–94, October 1976. 10.3

deman:04:ddp
[41] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49(11):2463–75, June 2004. 10.3

mazur:95:iar
[42] E. J. Mazur and R. Gordon. Interpolative algebraic reconstruction techniques without beam partitioning for computed tomography. *Med. Biol. Eng. Comput.*, 33(1):82–6, January 1995. 10.3

rossi:84:rfp
[43] D. J. Rossi and A. S. Willsky. Reconstruction from projections based on detection and estimation of objects—Parts I & II: Performance analysis and robustness analysis. *IEEE Trans. Acoust. Sp. Sig. Proc.*, 32(4):886–906, August 1984. 10.4

muller:90:mle
[44] S. P. Müller, M. F. Kijewski, S. C. Moore, and B. L. Holman. Maximum-likelihood estimation: a mathematical model for quantitation in nuclear medicine. *J. Nuc. Med.*, 31(10):1693–701, October 1990.

bichkar:98:tro
[45] R. S. Bichkar and A. K. Ray. Tomographic reconstruction of circular and elliptical objects using genetic algorithm. *IEEE Signal Proc. Letters*, 5(10):248–51, October 1998. 10.4

haacke:89:srt
[46] E. M. Haacke, Z. P. Liang, and S. H. Izen. Superresolution reconstruction through object modeling and parameter estimation. *IEEE Trans. Acoust. Sp. Sig. Proc.*, 37(4):592–5, April 1989. 10.4

haacke:89:cra
[47] E. M. Haacke, Z-P. Liang, and S. H. Izen. Constrained reconstruction: A superresolution, optimal signal-to-noise alternative to the Fourier transform in magnetic resonance imaging. *Med. Phys.*, 16(3):388–97, May 1989.

chiao:94:mbe-1
[48] P. C. Chiao, W. L. Rogers, N. H. Clinthorne, J. A. Fessler, and A. O. Hero. Model-based estimation for dynamic cardiac studies using ECT. *IEEE Trans. Med. Imag.*, 13(2):217–26, June 1994. 10.4

cunningham:00:4rf
[49] G. S. Cunningham and A. Lehovich. 4D reconstructions from low-count SPECT data using deformable models with smooth interior intensity variations. In *Proc. SPIE 3979 Medical Imaging 2000: Image Proc.*, pages 564–74, 2000. 10.4

battle:97:3tr
[50] X. L. Battle, G. S. Cunningham, and K. M. Hanson. 3D tomographic reconstruction using geometrical models. In *Proc. SPIE 3034 Med. Im. 1997: Im. Proc.*, pages 346–57, 1997. 10.4

battle:98:tru
[51] X. L. Battle, G. S. Cunningham, and K. M. Hanson. Tomographic reconstruction using 3D deformable models. *Phys. Med. Biol.*, 43(4):983–90, April 1998. 10.4

cunningham:98:tdr
[52] G. S. Cunningham, K. M. Hanson, and X. L. Battle. Three-dimensional reconstructions from low-count SPECT data using deformable models. *Optics Express*, 2(6):227–36, March 1998.

hanson:93:brb
[53] K. M. Hanson. Bayesian reconstruction based on flexible priors. *J. Opt. Soc. Am. A*, 10(5):997–1004, May 1993. 10.4

hanson:94:trb
[54] K. M. Hanson, G. S. Cunningham, and G. R. Jennings, Jr. D R Wolf. Tomographic reconstruction based on flexible geometric models. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 145–7, 1994. 10.4

hanson:97:uaf
[55] K. M. Hanson, G. S. Cunningham, and R. J. McKee. Uncertainty assessment for reconstructions based on deformable models. *Intl. J. Imaging Sys. and Tech.*, 8(6):506–12, 1997. 10.4

sitek:06:tru
[56] A. Sitek, R. H. Huesman, and G. T. Gullberg. Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud. *IEEE Trans. Med. Imag.*, 25(9):1172–9, September 2006. 10.4

benson:06:tdf
[57] T. M. Benson and J. Gregor. Three-dimensional focus of attention for iterative cone-beam micro-CT reconstruction. *Phys. Med. Biol.*, 51(18):4533–46, September 2006. 10.4

ziegler:08:iro
[58] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. *Med. Phys.*, 35(4):1317–27, April 2008. 10.4

gregor:11:ddp
[59] J. Gregor. Data-driven problem reduction for image reconstruction from projections using gift wrapping. *IEEE Trans. Nuc. Sci.*, 58(3):724–29, June 2011.

vidal:11:sc
[60] R. Vidal. Subspace clustering. *IEEE Sig. Proc. Mag.*, 28(2):52–68, March 2011. 10.6

balzano:12:ksw
[61] L. Balzano, A. Szlam, B. Recht, and R. Nowak. K-subspaces with missing data. In *IEEE Workshop on Statistical Signal Processing*, pages 612–5, 2012. 10.6

xie:12:mot
[62] Y. Xie, J. Huang, and R. Willett. Multiscale online tracking of manifolds. In *IEEE Workshop on Statistical Signal Processing*, pages 620–3, 2012. 10.6

adler:13:psc
[63] A. Adler, M. Elad, and Y. Hel-Or. Probabilistic subspace clustering via sparse representations. *IEEE Signal Proc. Letters*, 20(1):63–6, January 2013.

allard:12:msg
[64] W. K. Allard, G. Chen, and M. Maggioni. Multi-scale geometric methods for data sets II: Geometric multi-resolution analysis. *Applied and Computational Harmonic Analysis*, 32(3):435–62, May 2012.

elad:10
[65] M. Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, Berlin, 2010. 10.6

ravishankar:11:mir
[66] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imag.*, 30(5):1028–41, May 2011. 10.7, 10.10

xu:12:ldx
[67] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans. Med. Imag.*, 31(9):1682–97, September 2012.

aharon:06:ksa
[68] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Sig. Proc.*, 54(11):4311–22, November 2006. 10.8, 10.9

tropp:10:cmf
[69] J. A. Tropp and S. J. Wright. Computational methods for sparse solution of linear inverse problems. *Proc. IEEE*, 98(6):948–958, June 2010. 10.8

mallat:93:mpw
[70] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Sig. Proc.*, 41(12):3397–415, December 1993. 10.8

tropp:04:gig
[71] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Info. Theory*, 50(10):2231–42, October 2004. 10.8

mairal:10:olf
[72] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *J. Mach. Learning Res.*, 11:1960, January 2010. 10.9

gramfort:14:daf
[73] A. Gramfort, C. Poupon, and M. Descoteaux. Denoising and fast diffusion imaging with physically constrained sparse dictionary learning. *Med. Im. Anal.*, 18(1):36–49, January 2014. 10.9

smith:13:idl
[74] L. N. Smith and M. Elad. Improving dictionary learning: multiple dictionary updates and coefficient reuse. *IEEE Signal Proc. Letters*, 20(1):79–82, January 2013. 10.9

sadeghi:13:dlf
[75] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten. Dictionary learning for sparse representation: A novel approach. *IEEE Signal Proc. Letters*, 20(12):1195–8, December 2013. 10.9

cai::ddt
[76] J. F. Cai, S. Huang, H. Ji, Z. Shen, and G. B. Ye. Data-driven tight frame construction and image denoising, 2013.

elad:06:idv
[77] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Im. Proc.*, 15(12):3736–45, December 2006. 10.9, 10.10

bilgin:10:ddf
[78] A. Bilgin, Y. Kim, F. Liu, and M. S. Nadar. Dictionary design for compressed sensing MRI. In *Proc. Intl. Soc. Mag. Res. Med.*, page 4887, 2010. 10.9

doneva:10:csr
[79] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins. Compressed sensing reconstruction for magnetic resonance parameter mapping. *Mag. Res. Med.*, 64(4):1114–20, October 2010. 10.9

lu:12:fvi
[80] Y. Lu, J. Zhao, and G. Wang. Few-view image reconstruction with dual dictionaries. *Phys. Med. Biol.*, 57(1):173–190, January 2012. 10.9

yang:12:cdt
[81] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE Trans. Im. Proc.*, 21(8):3467–78, August 2012. 10.9

zhao:12:ddl
[82] B. Zhao, H. Ding, Y. Lu, G. Wang, J. Zhao, and S. Molloi. Dual-dictionary learning-based iterative reconstruction method in image reconstruction of spectral breast computed tomography. *Phys. Med. Biol.*, 2012. 10.9

marwah:13:clf
[83] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography - a new, high-resolution light field camera. In *SIGGRAPH*, 2013.

liu:13:alb
[84] Q. Liu, D. Liang, Y. Song, J. Luo, Y. Zhu, and W. Li. Augmented Lagrangian-based sparse representation method with dictionary updating for image deblurring. *SIAM J. Imaging Sci.*, 6(3):1689–718, 2013. 10.10

ravishankar:12:lst
[85] S. Ravishankar and Y. Bresler. Learning sparsifying transforms for image processing. In *Proc. IEEE Intl. Conf. on Image Processing*, pages 681–4, 2012. 10.10

ravishankar:12:lds
[86] S. Ravishankar and Y. Bresler. Learning doubly sparse transforms for image representation. In *Proc. IEEE Intl. Conf. on Image Processing*, pages 685–, 2012. 10.10

zhou:13:atf
[87] W. Zhou, J-F. Cai, and H. Gao. Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography. *Inverse Prob.*, 29(12):125006, December 2013.

elad:07:avs
[88] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Prob.*, 23(3):947–68, June 2007. 10.10

selesnick:09:srw
[89] I. W. Selesnick and Mário A T Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison

of analysis and synthesis priors. In *Proc. SPIE 7446 Wavelets XIII*, page 74460D, 2009. Wavelets XIII.