# A Unified Framework for Bases, Frames, Subspace Bases, and Subspace Frames

Gunnar Farnebäck
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

**Abstract**

Frame representations (e.g. wavelets) and subspace projections are important tools in many image processing applications. A unified framework for frames and subspace bases, as well as bases and subspace frames, is developed for finite dimensional vector spaces. Dual (subspace) bases and frames are constructed and the theory is generalized to weighted norms and seminorms. It is demonstrated how the framework applies to the cubic facet model, to normalized convolution, and to projection onto second degree polynomials.

**Keywords:** bases, frames, subspace bases, subspace frames, least squares, pseudo-inverse, dual vector sets, normalized convolution

## 1 Introduction

Frames and subspace bases, and of course bases, are well known concepts, which have been covered in several publications. Usually, however, they are treated as disparate entities. The idea behind this presentation of the material is to give a unified framework for bases, frames, and subspace bases, as well as the somewhat less known subspace frames. The basic idea is that the coefficients in the representation of a vector in terms of a frame, etc., can be described as solutions to various least squares problems. Using this to define what coefficients should be used, expressions for dual vector sets are derived. These results are then generalized to the case of weighted norms and finally also to the case of weighted seminorms. The presentation is restricted to finite dimensional vector spaces and relies heavily on matrix representations.

The background for the development of this framework is an increasing interest in foveally sampled images in the WITAS project [7]. The goal of the WITAS project is to develop an autonomous flying vehicle and to reduce the need for processing in the vision subsystem it would be advantageous to have a higher sampling density in areas of interest and lower elsewhere. In the analysis of such irregularly sampled images, e.g. motion estimation, the use of the frequency domain becomes quite complicated and therefore the attention has been turned to spatial domain methods. Subspace projection is a useful tool in these algorithms and the framework of this paper has been developed to lay a solid theoretical foundation for this kind of vector set representations.

Since many of the results in this work, especially those on least squares, are well known, it may be worth to point out what is new. The main contribution is of course the unification of the seemingly disparate concepts of frames and subspace bases in a least squares framework. Other things that seem to be novel is the simultaneous weighting in both the signal and the coefficient spaces for subspace frames, the full generalization of dual vector sets in section 4.4, and some of the results on seminorm weighted vector set representations in section 5.

The material of this paper can also be found in [3], together with applications to normalized convolution and to spatial domain methods for orientation and motion estimation. These methods are this far limited to regularly sampled images and adaptation to the irregularly sampled case is ongoing work. A short presentation of how this framework applies to the higher level tool normalized convolution is included in section 6 of this paper.

## 2 Preliminaries

To begin with, we review some basic concepts from (Numerical) Linear Algebra. All of these results are well known and can be found in any modern textbook on Numerical Linear Algebra, e.g. [4].

### 2.1 Notation

Let $\mathcal{C}^n$ be an $n$-dimensional complex vector space. Elements of this space are denoted by lower-case bold letters, e.g. $\mathbf{v}$, indicating $n \times 1$ column vectors. Upper-case bold letters, e.g. $\mathbf{F}$, denote complex matrices. With $\mathcal{C}^n$

is associated the standard inner product, $(\mathbf{f}, \mathbf{g}) = \mathbf{f}^*\mathbf{g}$, where $^*$ denotes conjugate transpose, and the Euclidian norm, $\|f\| = \sqrt{(\mathbf{f}, \mathbf{f})}$.

In this section $\mathbf{A}$ is an $n \times m$ complex matrix, $\mathbf{b} \in \mathcal{C}^n$, and $\mathbf{x} \in \mathcal{C}^m$.

## 2.2 The Linear Equation System

The linear equation system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

has a unique solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{2}$$

if and only if $\mathbf{A}$ is square and non-singular. If the equation system is overdetermined it does in general not have a solution and if it is underdetermined there are normally an infinite set of solutions. In these cases the equation system can be solved in a least squares and/or minimum norm sense, as discussed below.

## 2.3 The Linear Least Squares Problem

Assume that $n \geq m$ and that $\mathbf{A}$ is of rank $m$ (full column rank). Then the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ is not guaranteed to have a solution and the best we can do is to minimize the residual error.

The linear least squares problem

$$\arg \min_{\mathbf{x} \in \mathcal{C}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \tag{3}$$

has the unique solution

$$\mathbf{x} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b}. \tag{4}$$

If $\mathbf{A}$ is rank deficient the solution is not unique, a case which we return to in section 2.7.

## 2.4 The Minimum Norm Problem

Assume that $n \leq m$ and that $\mathbf{A}$ is of rank $n$ (full row rank). Then the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ may have more than one solution and to choose between them we take the one with minimum norm.

The minimum norm problem

$$\arg \min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{x}\|, \quad \mathcal{S} = \{\mathbf{x} \in \mathcal{C}^m; \mathbf{A}\mathbf{x} = \mathbf{b}\}. \tag{5}$$

has the unique solution

$$\mathbf{x} = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}\mathbf{b}. \tag{6}$$

If $\mathbf{A}$ is rank deficient it is possible that there is no solution at all, a case to which we return in section 2.7.

## 2.5 The Singular Value Decomposition

An arbitrary matrix $\mathbf{A}$ of rank $r$ can be factored by the Singular Value Decomposition, SVD, as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, $n \times n$ and $m \times m$ respectively. $\mathbf{\Sigma}$ is a diagonal $n \times m$ matrix

$$\mathbf{\Sigma} = \operatorname{diag}\left(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0\right), \tag{8}$$

where $\sigma_1, \ldots, \sigma_r$ are the non-zero singular values. The singular values are all real and $\sigma_1 \geq \ldots \geq \sigma_r > 0$. If $\mathbf{A}$ is of full rank we have $r = \min(n, m)$ and all singular values are non-zero.

## 2.6 The Pseudo-Inverse

The pseudo-inverse[1] $\mathbf{A}^\dagger$ of any matrix $\mathbf{A}$ can be defined via the SVD given by (7) and (8) as

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^*, \tag{9}$$

where $\mathbf{\Sigma}^\dagger$ is a diagonal $m \times n$ matrix

$$\mathbf{\Sigma}^\dagger = \operatorname{diag}\left(\tfrac{1}{\sigma_1}, \ldots, \tfrac{1}{\sigma_r}, 0, \ldots, 0\right). \tag{10}$$

We can notice that if $\mathbf{A}$ is of full rank and $n \geq m$, then the pseudo-inverse can also be computed as

$$\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* \tag{11}$$

and if instead $n \leq m$ then

$$\mathbf{A}^\dagger = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}. \tag{12}$$

If $m = n$ then $\mathbf{A}$ is quadratic and the condition of full rank becomes equivalent with non-singularity. It is obvious that both the equations (11) and (12) reduce to

$$\mathbf{A}^\dagger = \mathbf{A}^{-1} \tag{13}$$

in this case.

Regardless of rank conditions we have the following useful identities:

$$(\mathbf{A}^\dagger)^\dagger = \mathbf{A} \tag{14}$$

$$(\mathbf{A}^*)^\dagger = (\mathbf{A}^\dagger)^* \tag{15}$$

$$\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^\dagger\mathbf{A}^* \tag{16}$$

$$\mathbf{A}^\dagger = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^\dagger \tag{17}$$

[1] This pseudo-inverse is also known as the Moore-Penrose inverse.

## 2.7 The General Linear Least Squares Problem

The remaining case is when $\mathbf{A}$ is rank deficient. Then the equation $\mathbf{Ax} = \mathbf{b}$ is not guaranteed to have a solution and there may be more than one $\mathbf{x}$ minimizing the residual error. This problem can be solved as a simultaneous least squares and minimum norm problem.

The general (or rank deficient) linear least squares problem is stated as

$$\arg\min_{\mathbf{x}\in\mathcal{S}} \|\mathbf{x}\|, \quad \mathcal{S} = \{\mathbf{x} \in \mathcal{C}^m; \|\mathbf{Ax} - \mathbf{b}\| \text{ is min}\}, \quad (18)$$

i.e. among the least squares solutions, choose the one with minimum norm. Clearly this formulation contains both the ordinary linear least squares problem and the minimum norm problem as special cases. The unique solution is given in terms of the pseudo-inverse as

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} \quad (19)$$

Notice that by equations (11) – (13) this solution is consistent with (2), (4), and (6).

## 2.8 Numerical Aspects

Although the above results are most commonly found in books on *Numerical* Linear Algebra, only their algebraic properties are being discussed here. It should, however, be mentioned that e.g. equations (9) and (11) have numerical properties that differ significantly. The interested reader is referred to [1].

# 3 Representation by Sets of Vectors

If we have a set of vectors $\{\mathbf{f}_k\} \subset \mathcal{C}^n$ and wish to represent[2] an arbitrary vector $\mathbf{v}$ as a linear combination

$$\mathbf{v} \sim \sum c_k \mathbf{f}_k \quad (20)$$

of the given set, how should the coefficients $\{c_k\}$ be chosen? In general this question can be answered in terms of linear least squares problems.

## 3.1 Notation

With the set of vectors, $\{\mathbf{f}_k\}_{k=1}^m \subset \mathcal{C}^n$, is associated an $n \times m$ matrix

$$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m], \quad (21)$$

which effectively is a *reconstructing* operator because multiplication with an $m \times 1$ vector $\mathbf{c}$, $\mathbf{Fc}$, produces linear combinations of the vectors $\{\mathbf{f}_k\}$. In terms of the reconstruction matrix, equation (20) can be rewritten as

$$\mathbf{v} \sim \mathbf{Fc}, \quad (22)$$

where the coefficients $\{c_k\}$ have been collected in the vector $\mathbf{c}$.

The conjugate transpose of the reconstruction matrix, $\mathbf{F}^*$, gives an analyzing operator because $\mathbf{F}^*\mathbf{x}$ yields a vector containing the inner products between $\{\mathbf{f}_k\}$ and the vector $\mathbf{x} \in \mathcal{C}^n$.

[2] Ideally we would like to have equality in equation (20) but that cannot always be obtained.

Table 1: Definitions

| | | spans $\mathcal{C}^n$ | |
|---|---|---|---|
| | | yes | no |
| linearly | independent | basis | subspace basis |
| | dependent | frame | subspace frame |

## 3.2 Definitions

Let $\{\mathbf{f}_k\}$ be a subset of $\mathcal{C}^n$. If $\{\mathbf{f}_k\}$ spans $\mathcal{C}^n$ and is linearly independent it is called a *basis*. If it spans $\mathcal{C}^n$ but is linearly dependent it is called a *frame*. If it is linearly independent but does not span $\mathcal{C}^n$ it is called a *subspace basis*. Finally, if it neither spans $\mathcal{C}^n$, nor is linearly independent, it is called a *subspace frame*.[3] This relationship is depicted in table 1. If the properties of $\{\mathbf{f}_k\}$ are unknown or arbitrary we simply use *set of vectors* or *vector set* as a collective term.

## 3.3 Dual Vector Sets

We associate with a given vector set $\{\mathbf{f}_k\}$ the dual vector set $\{\tilde{\mathbf{f}}_k\}$, characterized by the condition that for an arbitrary vector $\mathbf{v}$ the coefficients $\{c_k\}$ in equation (20) are given as inner products between the dual vectors and $\mathbf{v}$,

$$c_k = (\tilde{\mathbf{f}}_k, \mathbf{v}) = \tilde{\mathbf{f}}_k^* \mathbf{v}. \quad (23)$$

This equation can be rewritten in terms of the reconstruction matrix $\tilde{\mathbf{F}}$ corresponding to $\{\tilde{\mathbf{f}}_k\}$ as

$$\mathbf{c} = \tilde{\mathbf{F}}^* \mathbf{v}. \quad (24)$$

The existence of the dual vector set is a nontrivial fact, which will be proved in the following sections for the various classes of vector sets.

## 3.4 Representation by a Basis

Let $\{\mathbf{f}_k\}$ be a basis. An arbitrary vector $\mathbf{v}$ can be written as a linear combination of the basis vectors, $\mathbf{v} = \mathbf{Fc}$, for a unique coefficient vector $\mathbf{c}$.[4]

[3] The notation used here is somewhat nonstandard. See section 3.9 for a discussion.

[4] The coefficients $\{c_k\}$ are of course also known as the *coordinates* for $\mathbf{v}$ with respect to the basis $\{\mathbf{f}_k\}$.

Because $\mathbf{F}$ is invertible in the case of a basis, we immediately get

$$\mathbf{c} = \mathbf{F}^{-1}\mathbf{v} \tag{25}$$

and it is clear from comparison with equation (24) that $\tilde{\mathbf{F}}$ exists and is given by

$$\tilde{\mathbf{F}} = (\mathbf{F}^{-1})^*. \tag{26}$$

In this very ideal case where the vector set is a basis, there is no need to state a least squares problem to find $\mathbf{c}$ or $\tilde{\mathbf{F}}$. That this could indeed be done is discussed in section 3.7.

### 3.5 Representation by a Frame

Let $\{\mathbf{f}_k\}$ be a frame. Because the frame spans $\mathcal{C}^n$, an arbitrary vector $\mathbf{v}$ can still be written as a linear combination of the frame vectors, $\mathbf{v} = \mathbf{F}\mathbf{c}$. This time, however, there are infinitely many coefficient vectors $\mathbf{c}$ satisfying the relation. To get a uniquely determined solution we add the requirement that $\mathbf{c}$ be of minimum norm. This is nothing but the minimum norm problem of section 2.4 and equation (6) gives the solution

$$\mathbf{c} = \mathbf{F}^*(\mathbf{F}\mathbf{F}^*)^{-1}\mathbf{v}. \tag{27}$$

Hence the dual frame exists and is given by

$$\tilde{\mathbf{F}} = (\mathbf{F}\mathbf{F}^*)^{-1}\mathbf{F}. \tag{28}$$

### 3.6 Representation by a Subspace Basis

Let $\{\mathbf{f}_k\}$ be a subspace basis. In general, an arbitrary vector $\mathbf{v}$ cannot be written as a linear combination of the subspace basis vectors, $\mathbf{v} = \mathbf{F}\mathbf{c}$. Equality only holds for vectors $\mathbf{v}$ in the subspace spanned by $\{\mathbf{f}_k\}$. Thus we have to settle for the $\mathbf{c}$ giving the closest vector $\mathbf{v}' = \mathbf{F}\mathbf{c}$ in the subspace. Since the subspace basis vectors are linearly independent we have the linear least squares problem of section 2.3 with the solution given by equation (4) as

$$\mathbf{c} = (\mathbf{F}^*\mathbf{F})^{-1}\mathbf{F}^*\mathbf{v}. \tag{29}$$

Hence the dual subspace basis exists and is given by

$$\tilde{\mathbf{F}} = \mathbf{F}(\mathbf{F}^*\mathbf{F})^{-1}. \tag{30}$$

Geometrically $\mathbf{v}'$ is the orthogonal projection of $\mathbf{v}$ onto the subspace.

### 3.7 Representation by a Subspace Frame

Let $\{\mathbf{f}_k\}$ be a subspace frame. In general, an arbitrary vector $\mathbf{v}$ cannot be written as a linear combination of the subspace frame vectors, $\mathbf{v} = \mathbf{F}\mathbf{c}$. Equality only holds for vectors $\mathbf{v}$ in the subspace spanned by $\{\mathbf{f}_k\}$. Thus we have to settle for a $\mathbf{c}$ giving the closest vector $\mathbf{v}' = \mathbf{F}\mathbf{c}$ in the subspace. Since the subspace frame vectors are linearly dependent there are also infinitely many $\mathbf{c}$ giving the same closest vector $\mathbf{v}'$, so to distinguish between these we choose the one with minimum norm. This is the general linear least squares problem of section 2.7 with the solution given by equation (19) as

$$\mathbf{c} = \mathbf{F}^\dagger\mathbf{v}. \tag{31}$$

Hence the dual subspace frame exists and is given by

$$\tilde{\mathbf{F}} = (\mathbf{F}^\dagger)^*. \tag{32}$$

The subspace frame case is the most general case since all the other ones can be considered as special cases. The only thing that happens to the general linear least squares problem formulated here is that sometimes there is an exact solution $\mathbf{v} = \mathbf{F}\mathbf{c}$, rendering the minimum residual error requirement superfluous, and sometimes there is a unique solution $\mathbf{c}$, rendering the minimum norm requirement superfluous. Consequently the solution given by equation (32) subsumes all the other ones, which is in agreement with equations (11) – (13).

### 3.8 The Double Dual

The dual of $\{\tilde{\mathbf{f}}_k\}$ can be computed from equation (32), applied twice, together with (14) and (15).

$$\tilde{\tilde{\mathbf{F}}} = \tilde{\mathbf{F}}^{\dagger *} = \mathbf{F}^{\dagger * \dagger *} = \mathbf{F}^{\dagger * * \dagger} = \mathbf{F}^{\dagger\dagger} = \mathbf{F}. \tag{33}$$

What this means is that if we know the inner products between $\mathbf{v}$ and $\{\mathbf{f}_k\}$ we can reconstruct $\mathbf{v}$ using the *dual* vectors. To summarize we have the two relations

$$\mathbf{v} \sim \mathbf{F}(\tilde{\mathbf{F}}^*\mathbf{v}) = \sum_k (\tilde{\mathbf{f}}_k, \mathbf{v})\mathbf{f}_k \quad \text{and} \tag{34}$$

$$\mathbf{v} \sim \tilde{\mathbf{F}}(\mathbf{F}^*\mathbf{v}) = \sum_k (\mathbf{f}_k, \mathbf{v})\tilde{\mathbf{f}}_k. \tag{35}$$

### 3.9 A Note on Notation

Usually a frame is defined by the frame condition,

$$A\|\mathbf{v}\|^2 \leq \sum_k |(\mathbf{f}_k, \mathbf{v})|^2 \leq B\|\mathbf{v}\|^2, \tag{36}$$

which must hold for some $A > 0$, some $B < \infty$, and all $\mathbf{v} \in \mathcal{C}^n$. In the finite dimensional setting used here the first inequality holds if and only if $\{\mathbf{f}_k\}$ spans all of $\mathcal{C}^n$ and the second inequality is a triviality as soon as the number of frame vectors is finite.

The difference between this definition and the one used in section 3.2 is that the bases are included in the set of frames. As we have seen that equation (28) is consistent with equation (26), the same convention could have been used here. The reason for not doing so is that the presentation would have become more involved.

Likewise, we may allow the subspace bases to span the whole $\mathcal{C}^n$, making bases a special case. Indeed, as has already been discussed to some extent, if subspace frames are allowed to be linearly independent, and/or span the whole $\mathcal{C}^n$, all the other cases can be considered special cases of subspace frames.

# 4 Weighted Norms

An interesting generalization of the theory developed so far is to exchange the Euclidian norms used in all minimizations for weighted norms.

## 4.1 Notation

Let the weighting matrix $\mathbf{W}$ be an $n \times n$ positive definite Hermitian matrix. The weighted inner product $(\cdot,\cdot)_{\mathbf{W}}$ on $\mathcal{C}^n$ is defined by

$$(\mathbf{f},\mathbf{g})_{\mathbf{W}} = (\mathbf{Wf},\mathbf{Wg}) = \mathbf{f}^*\mathbf{W}^*\mathbf{W}\mathbf{g} = \mathbf{f}^*\mathbf{W}^2\mathbf{g} \quad (37)$$

and the induced weighted norm $\|\cdot\|_{\mathbf{W}}$ is given by

$$\|\mathbf{f}\|_{\mathbf{W}} = \sqrt{(\mathbf{f},\mathbf{f})_{\mathbf{W}}} = \sqrt{(\mathbf{Wf},\mathbf{Wf})} = \|\mathbf{Wf}\|. \quad (38)$$

In this section $\mathbf{M}$ and $\mathbf{L}$ denote weighting matrices for $\mathcal{C}^n$ and $\mathcal{C}^m$ respectively. The notation from previous sections carry over unchanged.

## 4.2 The Weighted General Linear Least Squares Problem

The weighted version of the general linear least squares problem is stated as

$$\arg\min_{\mathbf{x}\in\mathcal{S}} \|\mathbf{x}\|_{\mathbf{L}}, \quad \mathcal{S} = \{\mathbf{x}\in\mathcal{C}^m; \|\mathbf{Ax}-\mathbf{b}\|_{\mathbf{M}} \text{ is min}\}. \quad (39)$$

This problem can be reduced to its unweighted counterpart by introducing $\mathbf{x}' = \mathbf{Lx}$, whereby equation (39) can be rewritten as

$$\begin{gathered}\arg\min_{\mathbf{x}'\in\mathcal{S}} \|\mathbf{x}'\|, \\ \mathcal{S} = \{\mathbf{x}'\in\mathcal{C}^m; \|\mathbf{MAL}^{-1}\mathbf{x}' - \mathbf{Mb}\| \text{ is min}\}.\end{gathered} \quad (40)$$

The solution is given by equation (19) as

$$\mathbf{x}' = (\mathbf{MAL}^{-1})^\dagger\mathbf{Mb}, \quad (41)$$

which after back-substitution yields

$$\mathbf{x} = \mathbf{L}^{-1}(\mathbf{MAL}^{-1})^\dagger\mathbf{Mb}. \quad (42)$$

## 4.3 Representation by Vector Sets

Let $\{\mathbf{f}_k\} \subset \mathcal{C}^n$ be any type of vector set. We want to represent an arbitrary vector $v \in \mathcal{C}^n$ as a linear combination of the given vectors,

$$\mathbf{v} \sim \mathbf{Fc}, \quad (43)$$

where the coefficient vector $\mathbf{c}$ is chosen so that

1. the distance between $\mathbf{v}' = \mathbf{Fc}$ and $\mathbf{v}$, $\|\mathbf{v}'-\mathbf{v}\|_{\mathbf{M}}$, is smallest possible, and
2. the length of $\mathbf{c}$, $\|\mathbf{c}\|_{\mathbf{L}}$, is minimized.

This is of course the weighted general linear least squares problem of the previous section, with the solution

$$\mathbf{c} = \mathbf{L}^{-1}(\mathbf{MFL}^{-1})^\dagger\mathbf{Mv}. \quad (44)$$

From the geometry of the problem one would suspect that $\mathbf{M}$ should not influence the solution in the case of a basis or a frame, because the vectors span the whole space so that $\mathbf{v}'$ equals $\mathbf{v}$ and the distance is zero, regardless of norm. Likewise $\mathbf{L}$ should not influence the solution in the case of a basis or a subspace basis. That this is correct can easily be seen by applying the identities (11) – (13) to the solution (44). In the case of a frame we get

$$\begin{aligned}\mathbf{c} &= \mathbf{L}^{-1}(\mathbf{MFL}^{-1})^*((\mathbf{MFL}^{-1})(\mathbf{MFL}^{-1})^*)^{-1}\mathbf{Mv} \\ &= \mathbf{L}^{-2}\mathbf{F}^*\mathbf{M}(\mathbf{MFL}^{-2}\mathbf{F}^*\mathbf{M})^{-1}\mathbf{Mv} \\ &= \mathbf{L}^{-2}\mathbf{F}^*(\mathbf{FL}^{-2}\mathbf{F}^*)^{-1}\mathbf{v},\end{aligned} \quad (45)$$

in the case of a subspace basis

$$\begin{aligned}\mathbf{c} &= \mathbf{L}^{-1}((\mathbf{MFL}^{-1})^*(\mathbf{MFL}^{-1}))^{-1}(\mathbf{MFL}^{-1})^*\mathbf{Mv} \\ &= \mathbf{L}^{-1}(\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{FL}^{-1})^{-1}\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{v} \\ &= (\mathbf{F}^*\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{v},\end{aligned} \quad (46)$$

and in the case of a basis

$$\mathbf{c} = \mathbf{L}^{-1}(\mathbf{MFL}^{-1})^{-1}\mathbf{Mv} = \mathbf{F}^{-1}\mathbf{v}. \quad (47)$$

## 4.4 Dual Vector Sets

It is not completely obvious how the concept of a dual vector set should be generalized to the weighted norm case. We would like to retain the symmetry relation from equation (33) and get correspondences to the representations (34) and (35). This can be accomplished by the weighted dual[5]

$$\tilde{\mathbf{F}} = \mathbf{M}^{-1}(\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M})^\dagger\mathbf{L}, \quad (48)$$

[5] To be more precise we should say ML-weighted dual, denoted $\tilde{\mathbf{F}}_{\mathbf{ML}}$. In the current context the extra index would only weigh down the notation, and has therefore been dropped.

which obeys the relations

$$\tilde{\tilde{\mathbf{F}}} = \mathbf{F}, \tag{49}$$

$$\mathbf{v} \sim \mathbf{F}\mathbf{L}^{-2}\tilde{\mathbf{F}}^*\mathbf{M}^2\mathbf{v}, \quad \text{and} \tag{50}$$

$$\mathbf{v} \sim \tilde{\mathbf{F}}\mathbf{L}^{-2}\mathbf{F}^*\mathbf{M}^2\mathbf{v}. \tag{51}$$

Unfortunately the two latter relations are not as easily interpreted as (34) and (35). The situation simplifies a lot in the special case where $\mathbf{L} = \mathbf{I}$. Then we have

$$\tilde{\mathbf{F}} = \mathbf{M}^{-1}(\mathbf{F}^*\mathbf{M})^\dagger, \tag{52}$$

which can be rewritten by identity (17) as

$$\tilde{\mathbf{F}} = \mathbf{F}(\mathbf{F}^*\mathbf{M}^2\mathbf{F})^\dagger. \tag{53}$$

The two relations (50) and (51) can now be rewritten as

$$\mathbf{v} \sim \mathbf{F}(\tilde{\mathbf{F}}^*\mathbf{M}^2\mathbf{v}) = \sum_k (\tilde{\mathbf{f}}_k, \mathbf{v})_\mathbf{M}\, \mathbf{f}_k, \quad \text{and} \tag{54}$$

$$\mathbf{v} \sim \tilde{\mathbf{F}}(\mathbf{F}^*\mathbf{M}^2\mathbf{v}) = \sum_k (\mathbf{f}_k, \mathbf{v})_\mathbf{M}\, \tilde{\mathbf{f}}_k. \tag{55}$$

Returning to the case of a general $\mathbf{L}$, the factor $\mathbf{L}^{-2}$ in (50) and (51) should be interpreted as a weighted linear combination, i.e. $\mathbf{F}\mathbf{L}^{-2}\mathbf{c}$ would be an $\mathbf{L}^{-1}$-weighted linear combination of the vectors $\{\mathbf{f}_k\}$, with the coefficients given by $\mathbf{c}$, analogously to $\mathbf{F}^*\mathbf{M}^2\mathbf{v}$ being the set of $\mathbf{M}$-weighted inner products between $\{\mathbf{f}_k\}$ and a vector $\mathbf{v}$.

# 5 Weighted Seminorms

The final level of generalization to be addressed here is when the weighting matrices are allowed to be semidefinite, turning the norms into seminorms. This has fundamental consequences for the geometry of the problem. The primary difference is that with a (proper) seminorm not only the vector $\mathbf{0}$ has length zero, but a whole subspace has. This fact has to be taken into account with respect to the terms spanning and linear dependence.[6]

## 5.1 The Seminorm Weighted General Linear Least Squares Problem

When $\mathbf{M}$ and $\mathbf{L}$ are allowed to be semidefinite[7] the solution to equation (39) is given by Eldén in [2] as

$$\mathbf{x} = (\mathbf{I} - (\mathbf{L}\mathbf{P})^\dagger\mathbf{L})(\mathbf{M}\mathbf{A})^\dagger\mathbf{M}\mathbf{b} + \mathbf{P}(\mathbf{I} - (\mathbf{L}\mathbf{P})^\dagger\mathbf{L}\mathbf{P})\mathbf{z}, \tag{56}$$

where $\mathbf{z}$ is arbitrary and $\mathbf{P}$ is the projection

$$\mathbf{P} = \mathbf{I} - (\mathbf{M}\mathbf{A})^\dagger\mathbf{M}\mathbf{A}. \tag{57}$$

Furthermore the solution is unique if and only if

$$\mathcal{N}(\mathbf{M}\mathbf{A}) \cap \mathcal{N}(\mathbf{L}) = \{\mathbf{0}\}, \tag{58}$$

where $\mathcal{N}(\cdot)$ denotes the null space. When there are multiple solutions, the first term of (56) gives the solution with minimum *Euclidian* norm.

If we make the restriction that only $\mathbf{M}$ may be semidefinite, the derivation in section 4.2 still holds and the solution is unique and given by equation (42) as

$$\mathbf{x} = \mathbf{L}^{-1}(\mathbf{M}\mathbf{A}\mathbf{L}^{-1})^\dagger\mathbf{M}\mathbf{b}. \tag{59}$$

## 5.2 Representation by Vector Sets and Dual Vector Sets

Here we have exactly the same representation problem as in section 4.3, except that that $\mathbf{M}$ and $\mathbf{L}$ may now be semidefinite. The consequence of $\mathbf{M}$ being semidefinite is that residual errors along some directions does not matter, while $\mathbf{L}$ being semidefinite means that certain linear combinations of the available vectors can be used for free. When both are semidefinite it may happen that some linear combinations can be used freely without affecting the residual error. This causes an ambiguity in the choice of the coefficients $\mathbf{c}$, which can be resolved by the additional requirement that among the solutions, $\mathbf{c}$ is chosen with minimum Euclidian norm. Then the solution is given by the first part of equation (56) as

$$\mathbf{c} = (\mathbf{I} - (\mathbf{L}(\mathbf{I} - (\mathbf{M}\mathbf{F})^\dagger\mathbf{M}\mathbf{F}))^\dagger\mathbf{L})(\mathbf{M}\mathbf{F})^\dagger\mathbf{M}\mathbf{v}. \tag{60}$$

Since this expression is something of a mess we are not going explore the possibilities of finding a dual vector set or analogues of the relations (50) and (51). Let us instead turn to the considerably simpler case where only $\mathbf{M}$ is allowed to be semidefinite. As noted in the previous section, we can now use the same solution as in the case with weighted norms, reducing the solution (60) to that given by equation (44),

$$\mathbf{c} = \mathbf{L}^{-1}(\mathbf{M}\mathbf{F}\mathbf{L}^{-1})^\dagger\mathbf{M}\mathbf{v}. \tag{61}$$

Unfortunately we can no longer define the dual vector set by means of equation (48), due to the occurence of an explicit inverse of $\mathbf{M}$. Applying identity (16) on (61), however, we get

$$\mathbf{c} = \mathbf{L}^{-1}(\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{F}\mathbf{L}^{-1})^\dagger\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{v} \tag{62}$$

and it follows that

$$\tilde{\mathbf{F}} = \mathbf{F}\mathbf{L}^{-1}(\mathbf{L}^{-1}\mathbf{F}^*\mathbf{M}^2\mathbf{F}\mathbf{L}^{-1})^\dagger\mathbf{L} \tag{63}$$

[6] Specifically, if a set of otherwise linearly independent vectors have a linear combination of norm zero, we say that they are *effectively* linearly dependent, since they for all practical purposes may as well have been.

[7] $\mathbf{M}$ and $\mathbf{L}$ may in fact be completely arbitrary matrices of compatible sizes.

yields a dual satisfying the relations (49) – (51). In the case that $\mathbf{L} = \mathbf{I}$ this expression simplifies further to (53), just as for weighted norms. For later reference we also notice that (61) reduces to

$$\mathbf{c} = (\mathbf{MF})^{\dagger}\mathbf{Mv}. \tag{64}$$

# 6 Applications

The theory developed above can be applied to many algorithms involving least squares fitting, subspace projections, or frame representations. For frames the use of dual frames is standard practice but with subspace bases it is not unusual that the coordinates are computed in a more complex way, involving the construction of an orthogonal basis, although the dual subspace basis would suffice. What is more interesting is that the generalizations to weighted norms and seminorms naturally can be applied to extend many algorithms.

## 6.1 Applying the Framework to Multidimensional Data

To apply this theory to signal processing problems, we notice that a discrete 1D signal of limited extent naturally can be treated as a finite dimensional vector, simply by collecting the sample values in a column. This means that we implicitly use a canonical "sample point" basis. For images and other multidimensional data we have exactly the same situation. Collecting the sample values, in some arbitrary but fixed order, we get a column vector with respect to a pixel basis. To give an example, the $2 \times 2$ image

$$\begin{matrix} 0 & 63 \\ 127 & 255 \end{matrix}$$

can be represented by the vector $\begin{pmatrix} 0 & 127 & 63 & 255 \end{pmatrix}^T$.

The limitation to discrete and finite signals is not much of a problem in practice since most algorithms for computational reasons do work on this kind of data. It should also be noticed that even if we should have signals of unlimited size, we are often interested in analyzing only a limited neighborhood of the signal at a time.

## 6.2 The Cubic Facet Model

In the cubic facet model [5], it is assumed that in each neighborhood of an image, the signal can be described by a cubic polynomial

$$\begin{aligned} f(x,y) = k_1 + k_2x + k_3y + k_4x^2 + k_5xy + k_6y^2 \\ + k_7x^3 + k_8x^2y + k_9xy^2 + k_{10}y^3. \end{aligned} \tag{65}$$

The coefficients $\{k_i\}$ are determined by a least squares fit within a square window of some size[8]. In [5] coefficients are first computed with respect to an orthogonal subspace basis and then transformed to the desired subspace basis. The orthogonal subspace basis is built by a Gram-Schmidt process in one dimension and a tensor product construction to get to two dimensions. Incidentally this construction yields a subspace basis of the form

$$\begin{aligned} \{1,\ x,\ y,\ x^2 - \alpha,\ xy,\ y^2 - \alpha, \\ x(y^2 - \alpha),\ y(x^2 - \alpha),\ (x^2 - \alpha)(y^2 - \alpha)\}, \end{aligned} \tag{66}$$

which does not really span the same subspace basis as the cubic polynomials.

With the framework from this paper it is straighforward to compute the dual subspace basis, with equation (30) or (32), so that the coefficients can be obtained directly by computing inner products. It is also straightforward to generalize the algorithm to use a weighted least squares fit, where the center pixels are considered more important than those farther away, simply by putting the weights into a diagonal weight matrix $\mathbf{M}$ and applying the theory from section 4. These ideas are explored in the following sections.

## 6.3 Normalized Convolution

Normalized convolution [6, 8] is a method for signal analysis that takes uncertainties in signal values into account and at the same time allows spatial localization of possibly unlimited analysis functions. Although a full description of the method would be outside the scope of this paper, we can still see how it relates to the framework developed here. First we need to establish a few terms in the context of the method

**Signal** The signal values in a neighborhood of a given point are represented by the vector $\mathbf{f}$.

**Certainty** Each signal value has a corresponding confidence value, represented in the neighborhood by the certainty vector $\mathbf{c}$. Reasons for uncertainty in the signal values include defective sensor elements, detected (but not corrected) transmission errors, varying confidence in the results from previous processing, and missing data outside the borders, i.e. edge effects. Certainty values are non-negative, with zero denoting missing data.

**Basis functions** The local signal model is given by a set of subspace basis vectors, $\{\mathbf{b}_i\}$, represented by the matrix $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m]$. The choice of basis functions is dependent on the application. Despite the name it is not really necessary that they are linearly independent.

[8] For simplicity we assume it to be at least $5 \times 5$.

**Applicability** The applicability $\mathbf{a}$ gives the relative importance of the points in the neighborhood, typically monotonically decreasing in all directions from the center point. Applicability values are non-negative. Points with zero applicability may as well be excluded from the neighborhood but can be kept if it is more convenient.

Let the diagonal matrices $\mathbf{W}_a = \operatorname{diag}(\mathbf{a})$, $\mathbf{W}_c = \operatorname{diag}(\mathbf{c})$, and $\mathbf{W}^2 = \mathbf{W}_a\mathbf{W}_c$. The operation of normalized convolution is at each point a question of representing a neighborhood of the signal, $\mathbf{f}$, by the set of vectors $\{\mathbf{b}_i\}$, using the weighted norm or seminorm $\|\cdot\|_\mathbf{W}$ in the signal space and the Euclidian norm in the coefficient space. The result of normalized convolution is at each point the set of coefficients $\mathbf{r}$ used in the vector set representation.

Alternatively we can state this in terms of a seminorm weighted general linear least squares problem

$$\arg\min_{\mathbf{r}\in\mathcal{S}} \|\mathbf{r}\|, \quad \mathcal{S} = \{\mathbf{r} \in \mathcal{C}^m; \|\mathbf{Br} - \mathbf{f}\|_\mathbf{W} \text{ is min}\}. \tag{67}$$

The solution is given by equation (64) as

$$\mathbf{r} = (\mathbf{WB})^\dagger \mathbf{Wf}. \tag{68}$$

For computational reasons it is often beneficial to expand the pseudo-inverse by identity (16), leading to

$$\begin{aligned} \mathbf{r} &= (\mathbf{B}^*\mathbf{W}^2\mathbf{B})^\dagger \mathbf{B}^*\mathbf{W}^2\mathbf{f} \\ &= (\mathbf{B}^*\mathbf{W}_a\mathbf{W}_c\mathbf{B})^\dagger \mathbf{B}^*\mathbf{W}_a\mathbf{W}_c\mathbf{f}, \end{aligned} \tag{69}$$

and if the basis functions actually constitute a subspace basis with respect to the seminorm $\mathbf{W}$, we can replace the pseudo-inverse in the previous equation with an ordinary inverse, yielding

$$\mathbf{r} = (\mathbf{B}^*\mathbf{W}_a\mathbf{W}_c\mathbf{B})^{-1} \mathbf{B}^*\mathbf{W}_a\mathbf{W}_c\mathbf{f}. \tag{70}$$

The inverse here is of course not computed explicitly since there are more efficient ways to solve a linear equation system. In the case that the certainty is constant over all neighborhoods, this can be further simplified to

$$\mathbf{r} = \tilde{\mathbf{B}}^*\mathbf{f}, \tag{71}$$

where the dual[9] $\tilde{\mathbf{B}}$ is given by

$$\tilde{\mathbf{B}} = \mathbf{W}_a\mathbf{B}(\mathbf{B}^*\mathbf{W}_a\mathbf{B})^{-1}. \tag{72}$$

If we compare the cubic facet model with normalized convolution, we can see that the former is incorporated as a special case of the latter, with cubic polynomials as basis functions, applicability identically one on a square, and certainty identically one. The generalization of the cubic facet model to a weighted least squares fit is included in normalized convolution by means of the applicability.

[9] Unfortunately this is not a proper dual as defined in section 4.4 but rather a quasi-dual.

An extensive presentation of normalized convolution, set in the framework developed in this paper, as well as applications to orientation and motion estimation, can be found in [3].

## 6.4 Projection onto Second Degree Polynomials

We now take a look at how normalized convolution can be applied to signal analysis by projection onto second degree polynomials. We start with the local signal model, expressed in a local coordinate system,

$$f(\mathbf{x}) \sim \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c, \tag{73}$$

where $\mathbf{A}$ is a symmetric matrix, $\mathbf{b}$ a vector and $c$ a scalar. The coefficients of the model can be estimated in terms of normalized convolution with the basis functions

$$\{1, x, y, z, x^2, y^2, z^2, xy, xz, yz\} \tag{74}$$

for the 3D case, with obvious generalizations to other dimensionalities. As will be discussed later, a Gaussian is a good choice of applicability. If we assume that we have constant certainty[10] we can use equation (72) to compute a dual basis so that the coefficients in the polynomial representation are computed by inner products with the dual basis vectors. It is useful to notice that these inner products can be computed over the whole signal as convolutions, with the reflected dual basis vectors as convolution kernels; a direct consequence of the definition of convolution.

To see the structure of the dual basis vectors we can rewrite equation (72) as

$$\begin{pmatrix} | & & | \\ \tilde{\mathbf{b}}_1 & \dots & \tilde{\mathbf{b}}_m \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{a}\cdot\mathbf{b}_1 & \dots & \mathbf{a}\cdot\mathbf{b}_m \\ | & & | \end{pmatrix} \mathbf{G}^{-1}, \tag{75}$$

where $\mathbf{G} = \mathbf{B}^*\mathbf{W}_a\mathbf{B}$ and $\cdot$ denotes pointwise multiplication. Hence we get the duals as linear combinations of the basis functions windowed by the applicability. The role of $\mathbf{G}^{-1}$ is to compensate for dependencies between the basis functions when they are not orthogonal, including non-orthogonalities caused by the windowing by $\mathbf{a}$.

With the polynomial basis and Gaussian applicability,[11] it can be shown that the structure of $\mathbf{G}^{-1}$ is particularly simple, so that the convolution kernels needed to compute $\mathbf{A}$ and $\mathbf{b}$ in equation (74) have the form

$$\begin{aligned} &x_i g(\mathbf{x}), \\ &(x_i^2 - \alpha) g(\mathbf{x}), \\ &x_i x_j g(\mathbf{x}), \quad i \neq j, \end{aligned} \tag{76}$$

[10] Typical assumption if we have no specific certainty information. Be aware though that it fails close to the borders of the signal.

[11] It does not have to be Gaussian but it must be sufficiently symmetric.

where $g(\mathbf{x})$ is the Gaussian applicability and $\alpha$ has exactly the value needed to eliminate the DC response.

One reason for the choice of a Gaussian applicability is that it is Cartesian separable, a property which is inherited by the convolution kernels (76). This means that $\mathbf{A}$ and $\mathbf{b}$ can be computed very efficiently solely by one-dimensional convolutions.

The second reason to choose a Gaussian applicability is that it is isotropic, i.e. rotation invariant. In [3] it is shown how the projections onto second degree polynomials can be used to estimate orientation. By evaluation on a simple 3D test volume, consisting of spherical shells and thus an even distribution of orientations, it turns out that an isotropic applicability is of utmost importance, especially in the absence of noise. With a Gaussian applicability the estimation is extremely accurate with a mean squared angular error as low as $0.11°$. A $5 \times 5 \times 5$ cube on the other hand gives an error of $13.5°$ and other tested applicabilities clearly show that (lack of) isotropy is by far the most significant factor. This is of particular interest considering that the naive approach of unweighted least squares fitting on a square or cube, cf. section 6.2, exactly corresponds to the use of the very anisotropic cube applicability.

It is interesting to compare this polynomial projection approach with signal analysis estimated from first and second derivatives. By the Maclaurin expansion, a sufficiently differentiable signal can be expanded in a neighborhood of the origin as

$$f(\mathbf{x}) = f(\mathbf{0}) + (\nabla f)^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H} \mathbf{x} + O(\|\mathbf{x}\|^3), \qquad (77)$$

where the gradient $\nabla f$ contains the first derivatives of $f$ at the origin and the Hessian $\mathbf{H}$ contains the second derivatives.

Clearly this expansion looks identical to the signal model (73) with $\mathbf{A} = \frac{1}{2}\mathbf{H}$, $\mathbf{b} = \nabla f$, and $c = f(0)$. As it happens, convolution with the kernels in (76) agrees exactly with estimation of the first and second derivatives of a signal convolved with a Gaussian. It should be stressed, however, that this relation is purely coincidental and depends on the special properties of the Gaussians. It does not hold for other choices of applicabilities.

From a conceptual point of view, subspace projection and differentiation are two very different operations. The latter is by definition an operation on an infinitesimal neighborhood and in order to be used with discretized signals it is necessary to perform some kind of smoothing, especially in the presence of noise. The former operation, on the other hand, is intended to approximate the signal over a larger neighborhood, specified by the applicability, and can very naturally be used with discretized signals.

# 7 Conclusions

It has been demonstrated that bases, frames, subspace bases, and subspace frames for finite dimensional vector spaces can be treated in a unified manner in a least squares framework. This framework allows generalizations to weighted norms and weighted seminorms. With the use of dual vector sets, representations by arbitrary sets of vectors become no more complicated or less efficient to compute than representations by orthonormal bases.

# Acknowledgments

The author wants to acknowledge the financial support of WITAS: The Wallenberg Laboratory for Information Technology and Autonomous Systems.

# References

[1] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Society for Industrial and Applied Mathematics, 1996.

[2] L. Eldén. A Weighted Pseudoinverse, Generalized Singular Values, and Constrained Least Squares Problems. *BIT*, 22:487–502, 1982.

[3] G. Farnebäck. Spatial Domain Methods for Orientation and Velocity Estimation. Lic. Thesis LiU-Tek-Lic-1999:13, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, March 1999. Thesis No. 755, ISBN 91-7219-441-3.

[4] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, second edition, 1989.

[5] R. M. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(1):58–68, January 1984.

[6] H. Knutsson and C-F. Westin. Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 515–523, New York City, USA, June 1993. IEEE.

[7] WITAS web page. `http://www.ida.liu.se/ext/witas/eng.html`.

[8] C-F. Westin. *A Tensor Framework for Multidimensional Signal Processing*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1994. Dissertation No 348, ISBN 91-7871-421-4.