Review of Optimal Recovery

D. Darian Muresan

Abstract

Optimal recovery is a tool that can be used with adaptively quadratic (AQua) image models to develop efficient interpolation algorithms. In this technical report optimal recovery is reviewed as it applies to AQua image models. The connection between optimal recovery, least squares and support vector machines is also slightly touched upon.

Keywords

image modeling, quadratic classes, interpolation, denoising, demosaicing

The theory of optimal recovery is detailed in [1] and [2]. Using the notation of [1] optimal recovery is reviewed, as it applies to the problem of image interpolation. The basic problem of image interpolation is that of approximating an unknown function $\mathbf{x}$ at pixel x_0 in terms of its known values at pixels $x_1, \ldots, x_k$, with the additional assumption that $\mathbf{x}$ is an element of a known linear space V. More generally, a linear functional $F(\mathbf{x})$ is estimated in terms of other known linear functionals $F_1(\mathbf{x}), \ldots, F_k(\mathbf{x})$. (A linear functional $F(\mathbf{x})$ can be any linear function of the elements $\mathbf{x}$: it can be the samples of $\mathbf{x}$, its derivatives, its integrals, etc.) Functionals F_i are also assumed to be linearly independent, or else one of the functional F_i can be written as a function of the others. If F is linearly dependent with F_i, then F can be determined exactly and there is no estimation problem. Therefore, the assumption is that $F, F_1, \ldots, F_k$ are all linearly independent.

The image is modeled as belonging to a certain ellipsoidal signal class K:

$$K = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{Q} \mathbf{x} \leq \epsilon \right\} \tag{1}$$

where $\mathbf{Q}$ is a symmetric positive definite matrix. (If matrix $\mathbf{Q}$ is symmetric positive semidefinite, it can be made symmetric positive definite by setting the zero singular values to a small, non-zero value. The final estimation results will not be affected much.) Note

Electrical and Computer Engineering department at Cornell University, Ithaca, NY. 14853

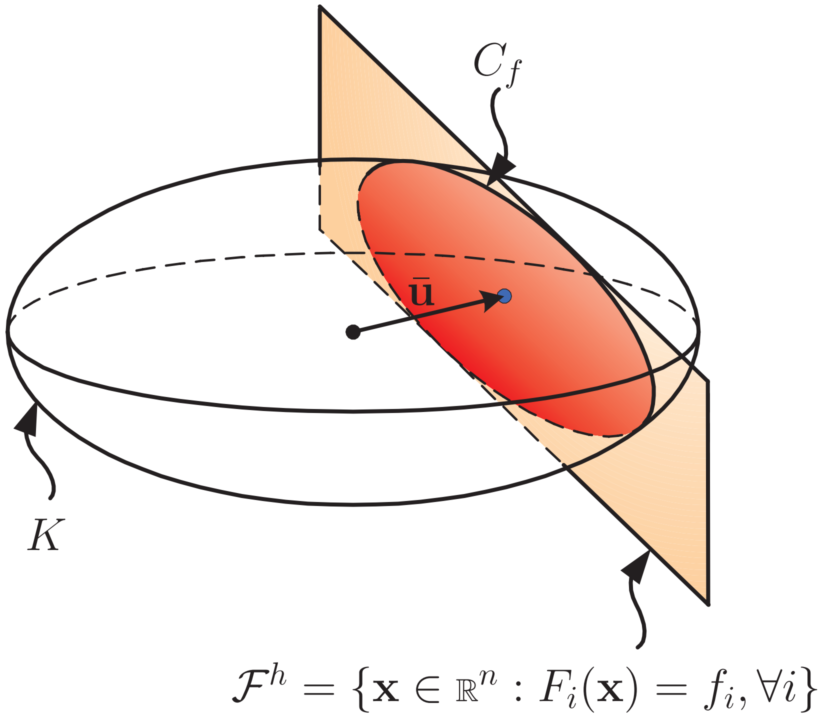

Fig. 1

INTERSECTION OF K WITH HYPER-PLANE $\mathcal{F}^h$.

the actual values of the functionals F_i by f_i. Then, the unknown function $\mathbf{x}$ lies in the set C_f:

$$C_f = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{Q}\mathbf{x} \leq \epsilon, F_i(\mathbf{x}) = f_i \text{ for } i = 1, \ldots, k\right\} \tag{2}$$

That is $\mathbf{x}$ lies in C_f, the hyper-circle defined by the intersection of the hyper-plane $\mathcal{F}^h = \{\mathbf{x} \in \mathbb{R}^n : F_i(\mathbf{x}) = f_i, \forall i\}$ with our ellipsoid signal class K, as shown in Fig. 1. Next, let $\bar{\mathbf{u}}$ be the minimum norm signal in C_f:

$$\|\bar{\mathbf{u}}\|_{\mathbf{Q}} = \inf_{F_i(\mathbf{x})=f_i} \|\mathbf{x}\|_{\mathbf{Q}} \tag{3}$$

and $\mathcal{F}$ be the subspaces parallel to the hyper-circle C_f:

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^n : F_1(\mathbf{x}) = \ldots = F_k(\mathbf{x}) = 0\}\,. \tag{4}$$

Our aim is to show that $F(\bar{\mathbf{u}})$ is the best approximation to the value of $F(\mathbf{x})$. That is

$F(\bar{\mathbf{u}})$ is the Chebyshev center [3] of $F(\mathbf{x})$ on C_f:

$$\sup_{\mathbf{x}\in C_f} |F(\bar{\mathbf{u}}) - F(\mathbf{x})| = \inf_{\mathbf{u}\in C_f} \sup_{\mathbf{x}\in C_f} |F(\mathbf{u}) - F(\mathbf{x})| \tag{5}$$

To begin with, we have the following theorem:

Theorem 1: The minimum norm signal $\bar{\mathbf{u}}$ is unique and orthogonal to subspace $\mathcal{F}$.

Proof: The fact that $\bar{\mathbf{u}}$ exists is obvious, since C_f is not empty. Uniqueness is proved by contradiction. If there exists $\bar{\mathbf{u}}_1$ and $\bar{\mathbf{u}}_2$ distinct minimum norm signals, then the non-zero vector $\mathbf{u} = (\bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_2) \in C_f$. In particular $\bar{\mathbf{u}}_1 + \frac{\mathbf{u}}{2} \in C_f$. By the triangle inequality

$$\|\bar{\mathbf{u}}_1 + \frac{\mathbf{u}}{2}\|_{\mathbf{Q}} \leq \|\bar{\mathbf{u}}_1\|_{\mathbf{Q}} + \|\frac{\mathbf{u}}{2}\|_{\mathbf{Q}} \tag{6}$$

$$< \|\bar{\mathbf{u}}_1\|_{\mathbf{Q}} \tag{7}$$

which is a contradiction since $\bar{\mathbf{u}}_1$ was assumed minimum norm. Thus $\bar{\mathbf{u}}$ is unique. To show orthogonality, we again proceed by contradiction. Assume that $\bar{\mathbf{u}}$ is not orthogonal to $\mathcal{F}$ and let $\mathbf{u}$ be the vector in C_f orthogonal to $\mathcal{F}$. Then $\bar{\mathbf{u}} = \mathbf{u} + (\bar{\mathbf{u}} - \mathbf{u})$. Since $(\bar{\mathbf{u}} - \mathbf{u}) \in C_f \Rightarrow \mathbf{u} \perp (\bar{\mathbf{u}} - \mathbf{u})$ and

$$\|\bar{\mathbf{u}}\|^2_{\mathbf{Q}} = \|\mathbf{u}\|^2_{\mathbf{Q}} + \|(\bar{\mathbf{u}} - \mathbf{u})\|^2_{\mathbf{Q}} \tag{8}$$

Unless $\bar{\mathbf{u}} = \mathbf{u}$, equation (8) implies that $\|\mathbf{u}\|_{\mathbf{Q}} < \|\bar{\mathbf{u}}\|_{\mathbf{Q}}$. This is a contradiction and $\bar{\mathbf{u}} \perp \mathcal{F}$
. Q.E.D.

Next, let's take a closer look at hyper-circle C_f, depicted in Fig. 2. Subspace $\mathcal{F}_0 = \{\mathbf{x} \in \mathbb{R}^n : F(\mathbf{x}) = 0, F_i(\mathbf{x}) = 0, \forall i\}$ is the subspace inside $\mathcal{F}$ for which the functional F is zero. Hyperspace $\mathcal{F}_0^h$ is any hyperspace in C_f, parallel with $\mathcal{F}_0$, for which the functional F is constant. Let $\bar{\mathbf{y}}$ be the unit norm element in $\mathcal{F}$ for which the functional F attains its least upper bound.

$$F(\bar{\mathbf{y}}) = \sup_{\mathbf{x}\in\mathcal{F}, \|\mathbf{x}\|_{\mathbf{Q}}=1} |F(\mathbf{x})|. \tag{9}$$

For any $\mathbf{x} \in C_f, F(\mathbf{x})$ is the functional of the projection of $\mathbf{x}$ onto $\bar{\mathbf{y}}$. This means $\bar{\mathbf{y}} \perp \mathcal{F}_0$ and more precisely $\bar{\mathbf{y}} \in \mathcal{F} \cap \mathcal{F}_0^{\perp}$. For any $\mathbf{x} \in C_f$ we can write:

$$\mathbf{x} = \bar{\mathbf{u}} + \mathbf{y} + \mathbf{w}, \tag{10}$$

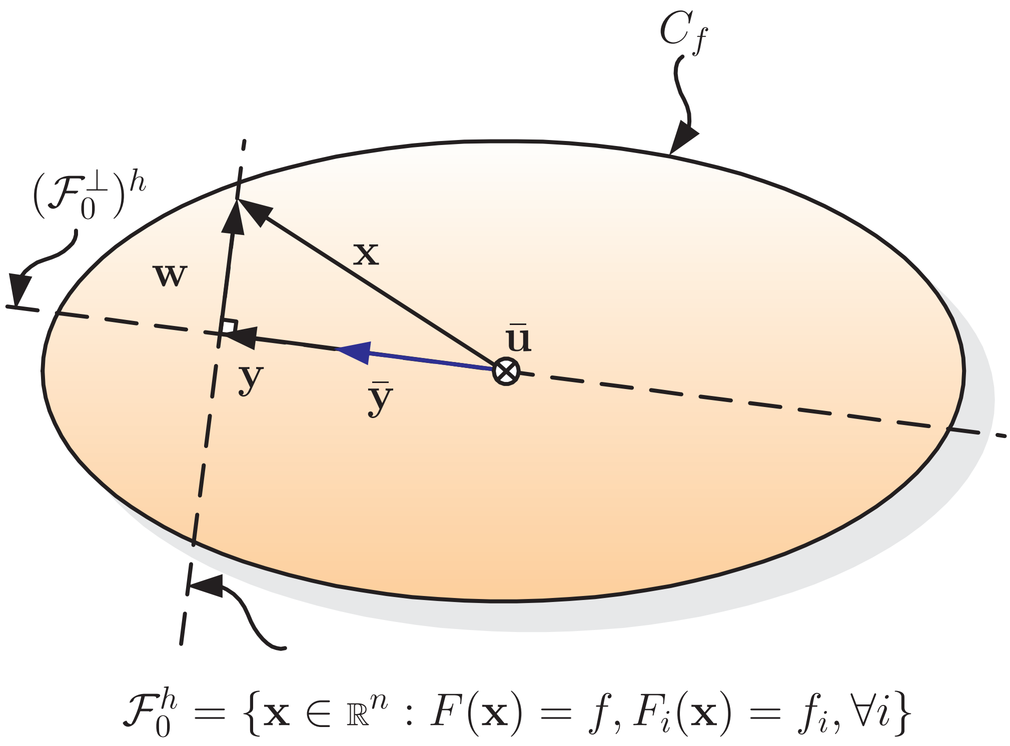

Fig. 2

HYPER-CIRCLE C_f.

where $\mathbf{y} = \alpha\bar{\mathbf{y}}$ (α constant) and $\mathbf{w} \in \mathcal{F}_0$, as shown in Fig. 2. To find α use equation (10) to obtain $F(\mathbf{y}) = F(\mathbf{x}) - F(\bar{\mathbf{u}}) - F(\mathbf{w}) = F(\mathbf{x}) - F(\bar{\mathbf{u}})$. This together with $F(\mathbf{y}) = \alpha F(\bar{\mathbf{y}})$ (since $\mathbf{y} = \alpha\bar{\mathbf{y}}$) gives:

$$\alpha = \frac{F(\mathbf{x}) - F(\bar{\mathbf{u}})}{F(\bar{\mathbf{y}})}$$

Equation (10) becomes:

$$\mathbf{x} = \bar{\mathbf{u}} + \frac{F(\mathbf{x}) - F(\bar{\mathbf{u}})}{F(\bar{\mathbf{y}})}\bar{\mathbf{y}} + \mathbf{w} \tag{11}$$

Vectors $\bar{\mathbf{u}}, \bar{\mathbf{y}}, \mathbf{w}$ are mutually orthogonal and from equation (11) we have

$$\|\mathbf{x}\|^2_{\mathbf{Q}} = \|\bar{\mathbf{u}}\|^2_{\mathbf{Q}} + \left(\frac{F(\mathbf{x}) - F(\bar{\mathbf{u}})}{F(\bar{\mathbf{y}})}\right)^2 + \|\mathbf{w}\|_{\mathbf{Q}} \tag{12}$$

$$\geq \|\bar{\mathbf{u}}\|_{\mathbf{Q}} + \left(\frac{F(\mathbf{x}) - F(\bar{\mathbf{u}})}{F(\bar{\mathbf{y}})}\right)^2 \tag{13}$$

for all $\mathbf{x} \in C_f$. Equation (13) becomes an equality when $\mathbf{w} = 0$, or when $\mathbf{x}$ lies in $\mathcal{F}_0^{\perp}$. For any $\mathbf{x}$ we have:

$$(F(\mathbf{x}) - F(\bar{\mathbf{u}}))^2 \leq F(\bar{\mathbf{y}})^2 \left(\|\mathbf{x}\|_{\mathbf{Q}} - \|\bar{\mathbf{u}}\|_{\mathbf{Q}}\right) \quad (14)$$

$$\leq F(\bar{\mathbf{y}})^2 \left(\epsilon - \|\bar{\mathbf{u}}\|_{\mathbf{Q}}\right) \quad (15)$$

Taking the square roots we obtain our bounds on the error of $F(\mathbf{x})$:

$$F(\bar{\mathbf{u}}) - F(\bar{\mathbf{y}}) \left(\epsilon - \|\bar{\mathbf{u}}\|_{\mathbf{Q}}\right)^{1/2} \leq F(\mathbf{x}) \leq F(\bar{\mathbf{u}}) + F(\bar{\mathbf{y}}) \left(\epsilon - \|\bar{\mathbf{u}}\|_{\mathbf{Q}}\right)^{1/2} \quad (16)$$

and these bounds are attained for the functions $\mathbf{x} \in C_f$:

$$\mathbf{x} = \bar{\mathbf{u}} \pm \left(\epsilon - \|\bar{\mathbf{u}}\|_{\mathbf{Q}}\right)^{1/2} \bar{\mathbf{y}} \quad (17)$$

which are the vectors at the intersection of the hyper-plane $(\mathcal{F}_0^{\perp})^h$ with the boundary of the hyper-circle C_f. Equation (16) also says that $F(\bar{\mathbf{u}})$ is the Chebyshev center of $F(\mathbf{x})$ on C_f.

The entire scenario of optimal recovery is captured in Fig. 2. Here are some observations in regards to Fig. 2. First, vector $\bar{\mathbf{u}}$ points orthogonally into the paper, in the center of C_f. Vectors $\mathbf{w}, \mathbf{y}, \bar{\mathbf{y}}$ and $\mathbf{x}$ live in $\mathcal{F}$ and are parallel with the vectors depicted in the figure. For ease of understanding, we drew them in the hyper-circle C_f.

Second, the error bound of equation (16) depends on the size of the hyper-circle C_f, which is directly proportional to ϵ. In other words, having a $\mathbf{Q}$ inner product and functionals F_i is not enough to find the error bounds, but it is enough to find $F(\bar{\mathbf{u}})$.

Third, $\bar{\mathbf{y}}$ does not have to point in the direction of the largest width of the hyper-circle C_f. Instead it is perpendicular to hyper-plane $\mathcal{F}_0^h$, a hyper-plane inside C_f for which the functional F is constant. It is for this reason that the bound on the error of $F(\mathbf{x})$ is dependent on the functional F, although the best vector estimate $\bar{\mathbf{u}}$ depends only on the functionals F_i and is independent of F. Since $\bar{\mathbf{u}}$ is independent of F it is consider an optimum estimate to the unknown function $\mathbf{x} \in C_f$.

A. *Finding $\bar{\mathbf{u}}$ and $\bar{\mathbf{y}}$*

Calculation of $\bar{\mathbf{u}}, F(\bar{\mathbf{u}}), \bar{\mathbf{y}}$, etc. is done using representors. By the Riesz representation theorem [4] there are elements $\phi, \phi_1, \ldots, \phi_k$ in $\mathbb{R}^n$ such that

$$F(\mathbf{x}) = (\phi, \mathbf{x})_{\mathbf{Q}}, \; F_i(\mathbf{x}) = (\phi_i, \mathbf{x})_{\mathbf{Q}}, \forall i \quad (18)$$

for all $\mathbf{x} \in K$. Vectors $\phi, \phi_1, \ldots, \phi_k$ are linearly independent since $F, F_1, \ldots, F_k$ are assumed to be linearly independent. Functionals $F_i(\mathbf{x}) = (\phi_i, \mathbf{x})_{\mathbf{Q}}$ remain constant for all $\mathbf{x} \in C_f$. That means subspace $\mathcal{F}$ is the set of all vectors in $\mathbb{R}^n$ orthogonal to the representors $\phi_i, \forall i$. Equivalently, $\phi_1, \ldots, \phi_k$ is a basis for $\mathcal{F}^\perp$. With $\bar{\mathbf{u}} \in \mathcal{F}^\perp$ it follows that $\bar{\mathbf{u}}$ is a linear combination of the representors ϕ_i. Similarly, $\mathcal{F}_0$ is the subspace of all vectors in $\mathbb{R}^n$ orthogonal to $\phi, \phi_1, \ldots, \phi_k$. Since $\bar{\mathbf{y}} \in \mathcal{F}_0^\perp$ it follows that $\bar{\mathbf{y}}$ is a linear combination of the representors $\phi, \phi_1, \ldots, \phi_k$. So here's what we have:

$$\bar{\mathbf{u}} = \sum_i c_i \phi_i \tag{19}$$

and

$$\bar{\mathbf{y}} = d\phi + \sum_i d_i \phi_i \tag{20}$$

We are left with finding the constants c_i and d, d_i. How do we find constants c_i? Since $\bar{u}$ must satisfy the given functionals we have

$$\begin{aligned} \underbrace{F_i(\bar{\mathbf{u}})}_{f_i} &= (\phi_i, \bar{\mathbf{u}})_{\mathbf{Q}} && (21) \\ &= \left(\phi_i, \sum_j c_j \phi_j\right)_{\mathbf{Q}} && (22) \end{aligned}$$

Written in matrix form this is equivalent to

$$\underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} (\phi_1, \phi_1)_{\mathbf{Q}} & (\phi_1, \phi_2)_{\mathbf{Q}} & \cdots & (\phi_1, \phi_k)_{\mathbf{Q}} \\ (\phi_1, \phi_2)_{\mathbf{Q}} & (\phi_2, \phi_2)_{\mathbf{Q}} & \cdots & (\phi_2, \phi_k)_{\mathbf{Q}} \\ \vdots & \vdots & \vdots & \vdots \\ (\phi_1, \phi_k)_{\mathbf{Q}} & \vdots & \vdots & (\phi_k, \phi_k)_{\mathbf{Q}} \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}}_{c} \tag{23}$$

Or,

$$c = G^{-1} b. \tag{24}$$

Matrix G is invertible since we assumed that all the representors are linearly independent. Similarly, we find d and d_i. Using the fact that $\bar{\mathbf{y}} \in \mathcal{F}$ it follows that $\bar{\mathbf{y}}$ is perpendicular to the representors $\phi_i, \forall i$:

$$d \begin{bmatrix} (\phi, \phi_1)_{\mathbf{Q}} \\ \vdots \\ (\phi, \phi_k)_{\mathbf{Q}} \end{bmatrix} + \begin{bmatrix} (\phi_1, \phi_1)_{\mathbf{Q}} & (\phi_1, \phi_2)_{\mathbf{Q}} & \cdots & (\phi_1, \phi_k)_{\mathbf{Q}} \\ (\phi_1, \phi_2)_{\mathbf{Q}} & (\phi_2, \phi_2)_{\mathbf{Q}} & \cdots & (\phi_2, \phi_k)_{\mathbf{Q}} \\ \vdots & \vdots & \vdots & \vdots \\ (\phi_1, \phi_k)_{\mathbf{Q}} & \vdots & \vdots & (\phi_k, \phi_k)_{\mathbf{Q}} \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_k \end{bmatrix} = 0 \tag{25}$$

From equation (25) we solve for d_i as a function of d. Vector $\bar{\mathbf{y}}$ from equation (20) is now defined as a function of only one unknown constant d. Constant d is found from the restriction that $\|\bar{y}\|_{\mathbf{Q}} = 1$.

Before finishing up this section, let's take a closer look at how we find the representors ϕ_i given functionals F_i and what implications this has on equations (24), (25) and the estimation problem. The functionals F_i are linear functionals of the input vector $\mathbf{x}$. In other words, there exists some vector $\mathbf{e}_i$ such that

$$\begin{aligned} F_i &= (\mathbf{x}, \mathbf{e}_i) && (26) \\ &= \mathbf{x}^T \mathbf{e}_i && (27) \end{aligned}$$

where the dot product in equation (27) is the standard inner product. For example, with $\mathbf{x} = [v_1\ v_2]^T \in \mathbb{R}^2$, $F_1 = v_1$ and $F_2 = v_1 + v_2$ then $\mathbf{e}_1 = [1\ 0]^T$ and $\mathbf{e}_2 = [1\ 1]^T$. Given the quadratic signal class K, the representors ϕ_i are found using

$$\begin{aligned} F_i &= (\mathbf{x}, \phi)_{\mathbf{Q}} && (28) \\ &= \mathbf{x}^T \mathbf{Q} \phi && (29) \end{aligned}$$

Substituting equation (29) into equation (27) we have

$$\phi_i = \mathbf{Q}^{-1} \mathbf{e}_i \tag{30}$$

and the $\mathbf{Q}$ inner product between two representors is

$$\begin{aligned} \left(\phi_i, \phi_j\right)_{\mathbf{Q}} &= \phi_i^T \mathbf{Q} \phi_j && (31) \\ &= \mathbf{e}_i^T \mathbf{Q}^{-T} \mathbf{Q} \mathbf{Q}^{-1} \mathbf{e}_j && (32) \\ &= \mathbf{e}_i^T \mathbf{Q}^{-1} \mathbf{e}_j && (33) \\ &= \mathbf{e}_i^t \phi_j && (34) \end{aligned}$$

From equations (30) and (33) one notices that in order to solve the estimation problem for $\bar{\mathbf{u}}$, $\bar{\mathbf{y}}$ and $F(\bar{\mathbf{u}})$ all we need is $\mathbf{Q}^{-1}$ and in fact $\mathbf{Q}$ is not used at all. This observation simplifies things significantly in case where $\mathbf{Q}$ is not invertible, but we can find $\mathbf{Q}^{-1}$ directly, without finding $\mathbf{Q}$ first, as is the case in future chapters.

I. Least Squares

The problem of least squares is a special case of optimal recovery. To understand the connection between least squares and optimal recovery, in this section we pose the least squares problem as a special case of the optimal recovery problem. Further, we show how solving the least squares problem forces a certain quadratic signal class K on our signal $\mathbf{x}$. To begin with, we start with the least squares problem: solve for $\mathbf{b}$, given

$$A\mathbf{b} = \mathbf{x} \tag{35}$$

with $A \in \mathbb{R}^{n \times m}, \mathbf{b} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$. Further, we assume that $n \geq m$ and that A is full rank, or that its column vectors are linearly independent.

If $\mathbf{x}$ is in the column space of A then equation (35) can be solved exactly. If $\mathbf{x}$ is not in the column space of A then the best we can do is to find the $\mathbf{b}$ for which $A\mathbf{b} = \hat{\mathbf{x}}$ is as close as possible to vector $\mathbf{x}$. In other words, we want to find $\mathbf{b}$ which minimizes the norm of the error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. By forcing the error to be orthogonal to the subspace spanned by the column vectors of A we obtain

$$\hat{\mathbf{x}} = A\left(A^T A\right)^{-1} A^T \mathbf{x} \tag{36}$$

and the norm squared of $\mathbf{e}$ is

$$\begin{aligned} \|\mathbf{e}\|^2 &= (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{e} && (37) \\ &= \mathbf{x}^T \mathbf{e} && (38) \\ &= \mathbf{x}^T \left[I - A\left(A^T A\right)^{-1} A^T\right] \mathbf{x} && (39) \end{aligned}$$

In other words, solving equation (35) is equivalent to saying that all the vectors which have error $\mathbf{e}$ after solving equation (35) leave in the quadratic signal class

$$K = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{Q} \mathbf{x} \leq \epsilon\right\} \tag{40}$$

with $\mathbf{Q} = I - A\left(A^T A\right)^{-1} A^T$.

Matrix $\mathbf{Q}$ has m eigenvalues equal to zero (the corresponding eigenvectors being in the range of A) and $n - m$ eigenvalues equal to one (the corresponding eigenvectors being in the complement of the range of A). When A is a square matrix, $\mathbf{Q} = 0$ and there is no error in solving equation (35). This is equivalent to saying that $K = \mathbb{R}^n$.

II. Support Vector Machines

This section is not intended to be a serious trip into the world of support vector machines, but rather a light discussion about the similarities between support vector machines, least squares and optimal recovery. For more serious discussions on support vector machines, see [5], [6], [7], [8].

The problem of support vector regression is the following optimization problem. Given a set of training data $\{(\mathbf{x}_1, y_1) \ldots (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^n \times \mathbb{R}$ find $\mathbf{w} \in \mathbb{R}^n, b \in \mathbb{R}$ such that

$$y_i = (\mathbf{w}, \mathbf{x}_i)_{\mathbf{K}} + b, \tag{41}$$

where $\mathbf{K}$ denotes the kernel inner product. (In the classification case, y_i can take values of only -1 and +1.) Assume for now that $\mathbf{K}$ is the simple inner product. Then, the optimization problem translates into finding a plane that best fits our training data. This is true, because for any dimension n, a plane can be described by its complement, which will always be a vector in $\mathbb{R}^n$. In equation (41), the complement of the plane is vector $\mathbf{w}$. Stating the problem this way, it is now obvious that the problem can be solved using regularized least squares. Of course, this is assuming that we want to minimize the sum of the square of the errors.

For ϵ insensitive loss functions, the problem can be stated as one of:

$$\text{minimize:} \quad \frac{1}{2}\|\mathbf{w}\|_{\mathbf{k}}^2 \tag{42}$$

$$\text{subject to:} \quad \begin{cases} y_i - (\mathbf{w}, \mathbf{x}_i)_{\mathbf{K}} - b & \leq \epsilon \\ -y_i + (\mathbf{w}, \mathbf{x}_i)_{\mathbf{K}} + b & \leq \epsilon \end{cases} \tag{43}$$

This resembles the optimal recovery problem, in that minimizing the norm of $\mathbf{w}$ is equivalent to minimizing the norm of $\bar{\mathbf{u}}$, subject to some linear constraints. And, the norm of $\mathbf{w}$

is dictated by the kernel we use, in a similar way the norm of $\bar{\mathbf{u}}$ is dictated by the matrix $\mathbf{Q}$.

In fact, a very challenging problem in support vector machines is how to choose the proper kernel for a given problem. This might be analogous with our image modeling problem of choosing $\mathbf{Q}$ adaptively from the image data.

References

[1] Michael Golomb and Hans F. Weinberger, *On Numerical Approximation*, R. E. Langer Ed. The University of Wisconsin Press, 1959.

[2] C. A. Michelli and T. J. Rivlin, *Optimal Estimation in Approximation Theory*, Eds. New York: Plenum 1976, 1976.

[3] R.G. Shenoy and T. W. Parks, "An optimal recovery approach to interpolation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1987–1996, 1992.

[4] Walter Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, Inc., 1976.

[5] Vladimir N. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, Inc., 1998.

[6] Alexander Smola, *Learning with Kernels*, Ph.D. thesis, Informatik der Technischen Universitat Berlin, Berlin, Germany, 1998.

[7] Bernhard Schölkopf, *Support Vector Learning*, Ph.D. thesis, Informatik der Technischen Universitat Berlin, Berlin, Germany, 1997.

[8] Bernhard Schölkopf, Christoper J. C. Burges, and Alexander J. Smola, *Advances in Kernel Methods*, The MIT Press, 1999.