

Partially Linear Estimation with Application to Image Deblurring Using Blurred/Noisy Image Pairs

Tomer Michaeli*, Daniel Sigalov, and Yonina Eldar

Technion – Israel Institute of Technology,
Haifa, 32000, Israel

{tomermic,dansigal}@tx.technion.ac.il, yonina@ee.technion.ac.il

Abstract. We address the problem of estimating a random vector X from two sets of measurements Y and Z , such that the estimator is linear in Y . We show that the partially linear minimum mean squared error (PLMMSE) estimator requires knowing only the second-order moments of X and Y , making it of potential interest in various applications. We demonstrate the utility of PLMMSE estimation in recovering a signal, which is sparse in a unitary dictionary, from noisy observations of it and of a filtered version of it. We apply the method to the problem of image enhancement from blurred/noisy image pairs. In this setting the PLMMSE estimator performs better than denoising or deblurring alone, compared to state-of-the-art algorithms. Its performance is slightly worse than joint denoising/deblurring methods, but it runs an order of magnitude faster.

Keywords: Bayesian estimation, minimum mean squared error, linear estimation.

1 Introduction

Bayesian estimation is concerned with the prediction of a random quantity X based on a set of observations Y , which are statistically related to X . It is well known that the estimator minimizing the mean squared error (MSE) is given by the conditional expectation $\hat{X} = \mathbb{E}[X|Y]$. There are various scenarios, however, in which the minimal MSE (MMSE) estimator cannot be used. This can either be due to implementation constraints, because of the fact that no closed form expression for $\mathbb{E}[X|Y]$ exists, or due to lack of complete knowledge of the joint distribution of X and Y . In these cases, one often resorts to linear estimation. The appeal of the linear MMSE (LMMSE) estimator is rooted in the fact that it possesses an easily implementable closed form expression, which merely requires knowledge of the joint first- and second-order moments of X and Y .

For example, the amount of computation required for calculating the MMSE estimate of a jump-Markov Gaussian random process from its noisy version

* This work was supported in part by a Google Research Award.

grows exponentially in time. By contrast, the LMMSE estimator in this setting possesses a simple recursive implementation, similar to the Kalman filter [1]. A similar problem arises in the area of sparse representations, in which the use of Bernoulli-Gaussian and Laplacian priors is very common. The complexity of calculating the MMSE estimator under the former prior is exponential in the vector's dimension, calling for approximate solutions [2, 3]. The MMSE estimator under the latter prior does not possess a closed form expression [4], which has motivated the use of alternative estimation strategies such as the maximum a-posteriori (MAP) method.

In practical situations, the reasons for not using the MMSE estimator may only apply to a subset of the measurements. Then, it may be desirable to construct an estimator that is linear in part of the measurements and nonlinear in the rest. One such scenario arises when estimating a sparsely representable vector X from two sets of measurements Y and Z , one blurred and one noisy. Indeed, as we show in this paper, when working with unitary dictionaries, the MMSE estimate $\mathbb{E}[X|Z]$ from the noisy measurements alone possesses an easy-to-implement closed form solution. However the complexity of computing the MMSE estimate $\mathbb{E}[X|Y, Z]$ from both sets of measurements is exponential. In this setting, the PLMMSE method, which is linear in Y , is computationally cheap and often comes close to the MMSE solution $\mathbb{E}[X|Y, Z]$ in terms of performance.

Partially linear estimation was studied in the statistical literature in the context of regression [5]. In this line of research, it is assumed that the conditional expectation $g(y, z) = \mathbb{E}[X|Y = y, Z = z]$ is linear in y . The goal, then, is to approximate $g(y, z)$ from a set of examples $\{x_i, y_i, z_i\}$ drawn independently from the joint distribution of X , Y and Z . In this paper, our goal is to derive the partially linear MMSE (PLMMSE) estimator. Namely, we do not make any assumptions on the structure of the MMSE estimate $\mathbb{E}[X|Y, Z]$, but rather look for the estimator that minimizes the MSE among all functions $g(Y, Z)$ that are linear in Y .

Due to space limitations, we state here the main results without their proofs, which can be found in [6].

2 Partially Linear Estimation

Suppose that X , Y and Z are random variables (RVs) taking values in \mathbb{R}^M , \mathbb{R}^N and \mathbb{R}^Q , respectively, such that X is the quantity to be estimated and Y and Z are two sets of measurements thereof. We denote by $\mathbf{\Gamma}_{XX}$, $\mathbf{\Gamma}_{XY}$, the auto-covariance of X and the cross-covariance of X and Y , respectively.

Our goal is to design a partially linear estimator of X based on Y and Z , which has the form

$$\hat{X} = \mathbf{A}Y + b(Z). \quad (1)$$

Here \mathbf{A} is a deterministic matrix and $b(z)$ is a vector-valued (Borel measurable) function.

Theorem 1. *The MMSE estimator of the form (1) is given by*

$$\hat{X} = \mathbf{\Gamma}_{XW} \mathbf{\Gamma}_{WW}^\dagger W + \mathbb{E}[X|Z], \quad (2)$$

where $W \triangleq Y - \mathbb{E}[Y|Z]$.

Note that (2) is of the form of (1) with $\mathbf{A} = \mathbf{\Gamma}_{XW} \mathbf{\Gamma}_{WW}^\dagger$ and $b(Z) = \mathbb{E}[X|Z] - \mathbf{\Gamma}_{XW} \mathbf{\Gamma}_{WW}^\dagger \mathbb{E}[Y|Z]$. As we show in [6], (2) can be equivalently written as

$$\hat{X} = \left(\mathbf{\Gamma}_{XY} - \mathbf{\Gamma}_{\hat{X}_Z \hat{Y}_Z} \right) \left(\mathbf{\Gamma}_{YY} - \mathbf{\Gamma}_{\hat{Y}_Z \hat{Y}_Z} \right)^\dagger \left(Y - \hat{Y}_Z \right) + \hat{X}_Z, \quad (3)$$

where $\hat{X}_Z \triangleq \mathbb{E}[X|Z]$ and $\hat{Y}_Z \triangleq \mathbb{E}[Y|Z]$. Therefore, all we need to know in order to be able to compute the PLMMSE estimator (2) is the covariance matrix $\mathbf{\Gamma}_{XY}$, the conditional expectation $\mathbb{E}[X|Z]$ and the joint distribution of Y and Z .

The intuition behind (2) is similar to that arising in dynamic estimation schemes, such as the Kalman filter. Specifically, we begin by constructing the MMSE estimate $\mathbb{E}[X|Z]$ of X from Z . We then update it with the LMMSE estimate of X based on the *innovation* W of Y with respect to $\mathbb{E}[X|Z]$.

One particularly interesting example is the case where X is observed through two linear systems as

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} \mathbf{H} \\ \mathbf{G} \end{pmatrix} X + \begin{pmatrix} U \\ V \end{pmatrix}, \quad (4)$$

where U and V are statistically independent. It is easily shown that in this setting, the PLMMSE estimate reduces to

$$\hat{X} = \mathbf{A}Y + (\mathbf{I} - \mathbf{H}\mathbf{A})\hat{X}_Z, \quad (5)$$

where \mathbf{I} denotes the identity matrix and

$$\mathbf{A} = \mathbf{C}\mathbf{H}^T (\mathbf{H}\mathbf{C}\mathbf{H}^T + \mathbf{\Gamma}_{UU})^\dagger \quad (6)$$

with $\mathbf{C} = \mathbf{\Gamma}_{XX} - \mathbf{\Gamma}_{\hat{X}_Z \hat{X}_Z}$.

3 Application to Sparse Approximations

Consider the situation in which X is known to be sparsely representable in a unitary dictionary $\mathbf{\Psi} \in \mathbb{R}^{M \times M}$ in the sense that

$$X = \mathbf{\Psi}A \quad (7)$$

for some RV A that is sparse with high probability. More concretely, we assume, as in [2, 3], a Bernoulli-Gaussian prior, so that the elements of A are given by

$$A_i = S_i B_i, \quad i = 1, \dots, M, \quad (8)$$

where the RVs $\{B_i\}$ and $\{S_i\}$ are statistically independent, $B_i \sim \mathcal{N}(0, \sigma_{B_i}^2)$ and $\mathbb{P}(S_i = 1) = 1 - \mathbb{P}(S_i = 0) = p_i$.

Assume X is observed through two linear systems, as in (4), where \mathbf{H} is an arbitrary matrix, \mathbf{G} is an orthogonal matrix satisfying $\mathbf{G}^T \mathbf{G} = \alpha^2 \mathbf{I}$ for some $\alpha \neq 0$, and U and V are Gaussian RVs with $\mathbf{\Gamma}_{UU} = \sigma_U^2 \mathbf{I}$ and $\mathbf{\Gamma}_{VV} = \sigma_V^2 \mathbf{I}$. In this case the expression for the MMSE estimate $\mathbb{E}[X|Y, Z]$ comprises 2^M summands [2] rendering its computation prohibitively expensive even for modest values of M . Various approaches have been devised to approximate this solution by a small number of terms (see *e.g.*, [2, 3] and references therein).

There are some special cases, however, in which the MMSE estimate possesses a simple structure, which can be implemented efficiently. One such case is when both the channel's response and the dictionary over which X is sparse correspond to orthogonal matrices. As in our setting $\mathbf{\Psi}$ is unitary and \mathbf{G} is orthogonal, this implies that we can efficiently compute the MMSE estimate $\mathbb{E}[X|Z]$ of X from Z . Therefore, instead of resorting to schemes for approximating $\mathbb{E}[X|Y, Z]$, we can employ the PLMMSE estimator of X based on Y and Z , which, in this situation, possesses the simple closed form expression (5). This approach is particularly effective when the SNR of the observation Y is much worse than that of Z , since the MMSE estimate $\mathbb{E}[X|Y, Z]$ in this case is close to being partially linear in Y . Such a setting is demonstrated in the sequel. We have the following result.

Theorem 2. *The MMSE estimate of X of (7) given Z of (4) is*

$$\mathbb{E}[X|Z] = \mathbf{\Psi} \tilde{f} \left(\frac{1}{\alpha} \mathbf{\Psi}^T \mathbf{G}^T Z \right), \quad (9)$$

where $\tilde{f}(\tilde{z}) = (f(\tilde{z}_1), \dots, f(\tilde{z}_M))^T$, with

$$f(\tilde{z}_i) = \frac{\frac{\alpha \sigma_{B_i}^2}{\alpha^2 \sigma_{B_i}^2 + \sigma_V^2} p_i \mathcal{N}(\tilde{z}_i; 0, \alpha^2 \sigma_{B_i}^2 + \sigma_V^2) \tilde{z}_i}{p_i \mathcal{N}(\tilde{z}_i; 0, \alpha^2 \sigma_{B_i}^2 + \sigma_V^2) + (1 - p_i) \mathcal{N}(\tilde{z}_i; 0, \sigma_V^2)}. \quad (10)$$

Here, $\mathcal{N}(\alpha; \mu, \sigma^2)$ denotes the normal probability density function with mean μ and variance σ^2 , evaluated at α .

Therefore, if, *e.g.*, $\mathbf{\Psi}$ is a wavelet basis and $\mathbf{G} = \mathbf{I}$ (so that $\alpha = 1$), then $\mathbb{E}[X|Z]$ can be efficiently computed by taking the wavelet transform of Z (multiplication by $\mathbf{\Psi}^T$), applying a scalar shrinkage function on each of the coefficients (namely calculating $f(\tilde{z}_i)$ for the i th coefficient) and applying the inverse wavelet transform (multiplication by $\mathbf{\Psi}$) on the result.

Equipped with a closed form expression for $\mathbb{E}[X|Z]$, we can now compute the terms needed for implementing the PLMMSE estimator (5). First, we note that

$$\mathbf{\Gamma}_{XX} = \mathbf{\Psi} \mathbf{\Gamma}_{AA} \mathbf{\Psi}^T, \quad (11)$$

where $\mathbf{\Gamma}_{AA}$ is a diagonal matrix with $(\mathbf{\Gamma}_{AA})_{i,i} = p_i \sigma_{B_i}^2$. Similarly,

$$\mathbf{\Gamma}_{\tilde{X}_Z \tilde{X}_Z} = \mathbf{\Psi} \text{Cov}(\tilde{f}(\tilde{Z})) \mathbf{\Psi}^T, \quad (12)$$

where $\text{Cov}(\tilde{f}(\tilde{Z}))$ is a diagonal matrix whose (i, i) element is $\beta_i = \text{Var}(f(\tilde{Z}_i))$. This is due to the fact that the elements of \tilde{Z} are statistically independent and

the fact that the function $\tilde{f}(\cdot)$ operates element-wise on its argument. Hence, the PLMMSE estimator is given by (5) with $\mathbb{E}[X|Z]$ of (9) and with the matrix

$$\mathbf{A} = \mathbf{\Psi} \mathbf{C} \mathbf{\Psi}^T \mathbf{H}^T \left(\mathbf{H} \mathbf{\Psi} \mathbf{C} \mathbf{\Psi}^T \mathbf{H}^T + \sigma_U^2 \mathbf{I} \right)^\dagger, \quad (13)$$

where here $\mathbf{C} = \mathbf{\Gamma}_{AA} - \text{Cov}(\tilde{f}(\tilde{Z})) = \text{diag}(p_1 \sigma_{B_1}^2 - \beta_1, \dots, p_M \sigma_{B_M}^2 - \beta_M)$. Observe that there is generally no closed form expression for the scalars β_i , rendering it necessary to compute them numerically.

An important special case corresponds to the setting in which $p_i = p$ and $\sigma_{B_i}^2 = \sigma_B^2$ for every i . In this situation, we also have that $\beta_i = \beta$ for every i . Furthermore,

$$\mathbf{\Gamma}_{XX} = \mathbf{\Psi} (p \sigma_B^2 \mathbf{I}) \mathbf{\Psi}^T = p \sigma_B^2 \mathbf{I} \quad (14)$$

and

$$\mathbf{\Gamma}_{\hat{X}_Z \hat{X}_Z} = \mathbf{\Psi} (\beta \mathbf{I}) \mathbf{\Psi}^T = \beta \mathbf{I}, \quad (15)$$

so that \mathbf{A} is simplified to

$$\mathbf{A} = (p \sigma_B^2 - \beta) \mathbf{H}^T \left((p \sigma_B^2 - \beta) \mathbf{H} \mathbf{H}^T + \sigma_U^2 \mathbf{I} \right)^\dagger. \quad (16)$$

As can be seen, here \mathbf{A} does not involve multiplication by $\mathbf{\Psi}$ or $\mathbf{\Psi}^T$. Thus, if \mathbf{H} corresponds to a convolution operation, so does \mathbf{A} , meaning that it can be efficiently applied in the Fourier domain.

3.1 Image Deblurring with Blurred/Noisy Image Pairs

When taking photos in dim light using a hand-held camera, there is a tradeoff between noise and motion blur, which can be controlled by tuning the shutter speed. Using a long exposure time, the image typically comes out blurred due to camera shake. On the other hand, with a short exposure time (and high camera gain), the image is very noisy. In [7] it was demonstrated how a high quality image can be constructed by properly processing two images of the same scene, one blurred and one noisy.

We now show how the PLMMSE approach can be applied in this setting to obtain plausible recoveries at a speed several orders of magnitude faster than any other sparsity-based method. In our setting X , Y and Z correspond, respectively, to the original, blurred (and slightly noisy) and noisy images. Thus, the measurement model is that described by (4), where \mathbf{H} corresponds to spatial convolution with some blur kernel, $\mathbf{G} = \mathbf{I}$, and U and V correspond to white Gaussian noise images with small and large variances respectively. We further assume that the image X is sparse in some orthogonal wavelet basis $\mathbf{\Psi}$, such that it can be written as in (7) and (8).

As we have seen, in this setting, the PLMMSE estimator can be computed in two stages. First, we calculate $\hat{X}_Z = \mathbb{E}[X|Z]$ by computing the wavelet transform $\tilde{Z} = \mathbf{\Psi}^T Z$, applying the scalar shrinkage function (10) on each wavelet

coefficient, and taking the inverse wavelet transform of the result. This stage requires knowledge of the parameters $\{p_i\}$, $\{\sigma_{B_i}^2\}$ and σ_V^2 . To this end, we assume that p_i and $\sigma_{B_i}^2$ are the same for wavelets coefficients at the same level. Namely, all wavelet coefficients of Z at level ℓ correspond to independent draws from the Gaussian mixture

$$f_{\tilde{z}_i}(\tilde{z}) = p^\ell \mathcal{N}(\tilde{z}; 0, \alpha^2 \sigma_{B^\ell}^2 + \sigma_V^2) + (1 - p) \mathcal{N}(\tilde{z}; 0, \sigma_V^2). \quad (17)$$

Consequently, p^ℓ , $\sigma_{B^\ell}^2$ and σ_V^2 can be estimated by expectation maximization (EM). In our experiments, we assumed that σ_V^2 is known.

In the second stage, the denoised image \hat{X}_Z needs to be combined with the blurred image Y using (5) with \mathbf{A} of (13). As discussed in Section 3, this can be carried out very efficiently if $p_i = p$ and $\sigma_{B_i}^2 = \sigma_B^2$ for all i . For the sake of efficiency we therefore abandon the assumption that p_i and $\sigma_{B_i}^2$ vary across wavelet levels and assume henceforth that all wavelet coefficients are independent and identically distributed. In this case, \mathbf{A} corresponds to the filter

$$A(\omega) = \frac{(\sigma_A^2 - \beta)H^*(\omega)}{(\sigma_A^2 - \beta)|H(\omega)|^2 + \sigma_U^2}, \quad (18)$$

where $H(\omega)$ is the frequency response of the blur kernel. Consequently, the final PLMMSE estimate corresponds to the inverse Fourier transform of

$$\hat{X}_{\text{PLMMSE}}^{\text{F}}(\omega) = \frac{(\sigma_A^2 - \beta)H^*(\omega)Y^{\text{F}}(\omega) + \sigma_U^2 \hat{X}_Z^{\text{F}}(\omega)}{(\sigma_A^2 - \beta)|H(\omega)|^2 + \sigma_U^2}, \quad (19)$$

where $Y^{\text{F}}(\omega)$ and $\hat{X}_Z^{\text{F}}(\omega)$ denote the Fourier transforms of Y and \hat{X}_Z , respectively. In our experiment, we assumed that the blur $H(\omega)$ and noise variance σ_U^2 are known. In practice, they can be estimated from Y and Z , as proposed in [7]. This stage also requires knowing the scalars $\sigma_A^2 = \mathbb{E}[A^2]$ and $\beta = \mathbb{E}[f^2(\tilde{z})]$, which we estimate as $\hat{\sigma}_A^2 = \frac{1}{M} \sum_{i=1}^M \tilde{z}_i^2 - \sigma_V^2$ and $\hat{\beta} = \frac{1}{M} \sum_{i=1}^M f^2(\tilde{z}_i)$.

Fig. 1 demonstrates our approach on the 512×512 Gold-hill image. In this experiment, the blur corresponded to a Gaussian kernel with standard deviation 3.2. To model a situation in which the noise in Y is due only to quantization errors, we chose $\sigma_U = 1/\sqrt{12} \approx 0.3$ and $\sigma_V = 45$. These parameters correspond to a peak signal to noise ratio (PSNR) of 25.08dB for the blurred image and 15.07dB for the noisy image.

We used the orthogonal Symlet wavelet of order 4 and employed 10 EM iterations to estimate p^ℓ and $\sigma_{B^\ell}^2$ in each wavelet level. The entire process takes 1.1 seconds on a Dual-Core 3GHz computer with un-optimized Matlab code. We note that our approach can be viewed as a smart combination of Wiener filtering for image deblurring and wavelet thresholding for image denoising, which are among the simplest and fastest methods available. Consequently, the running time is at least an order of magnitude faster than any other sparsity-based methods (see, *e.g.*, comparisons in [2]).

As can be seen in Fig. 1, the quality of the recoveries corresponding to the denoised image \hat{X}_Z and deblurred image \hat{X}_Y^{L} is rather poor with respect to the

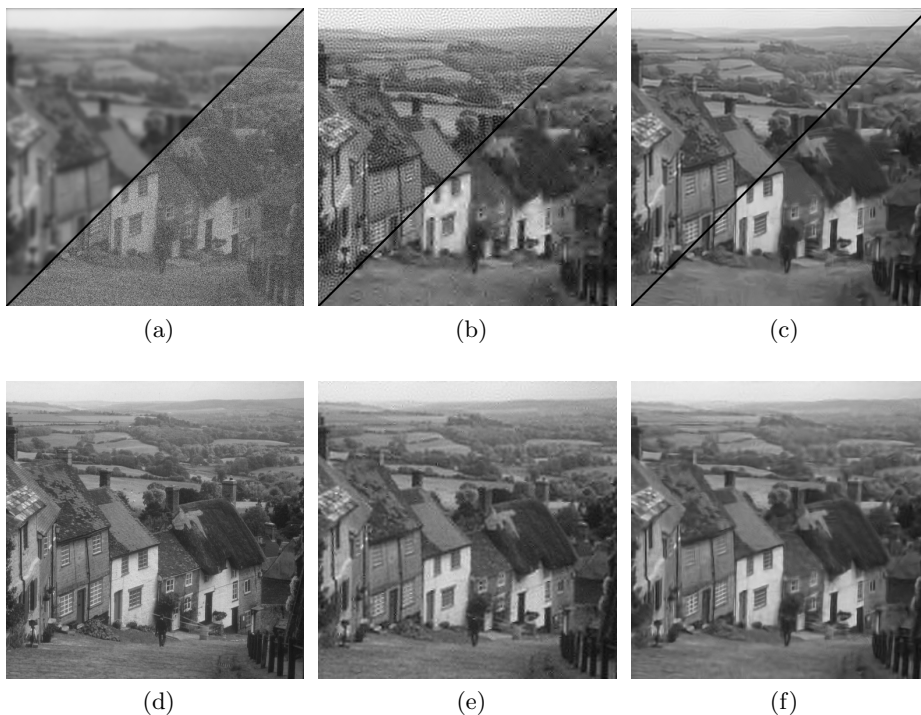


Fig. 1: Deblurring with a blurred/noisy image pair using PLMMSE estimation and RD [7]. (a) Blurred image Y (top left) and noisy image Z (bottom-right). (b) LMMSE-deblurred image \hat{X}_Y^L (top-left) and MMSE-denoised image \hat{X}_Z (bottom-right). (c) BM3D-deblurred image (top left) and BM3D-denoised image (bottom-right). (d) Original image X . (e) PLMMSE estimate \hat{X}_{PLMMSE} from Y and Z . (f) RD recovery.

state-of-the-art BM3D denoising method [8] and BM3D deblurring algorithm [9]. However, the quality of the joint estimate \hat{X}_{PLMMSE} surpasses each of these techniques. The residual deconvolution (RD) method [7] for joint deblurring and denoising outperforms the PLMMSE method in terms of recovery error but the visual differences are not prominent.

A quantitative comparison on several test images is given in Table 1. The PSNR attained by the PLMMSE method is, on average, 0.3dB higher than BM3D deblurring, 0.4db higher than BM3D denoising, and 0.8dB lower than RD. In terms of running times, however, our method is, on average, 11 times faster than BM3D deblurring, 16 times faster than BM3D denoising and 18 times faster than RD. Note that RD requires initialization with a denoised version of Z , for which purpose we used the BM3D algorithm. Hence, the running times

Table 1: Performance of deblurring/denoising on several images.

	\hat{X}_Z	\hat{X}_Y^L	BM3D Denoise	BM3D Deblur	PLMMSE	RD
Boat	25.39/0.83	23.45/0.06	27.85/13.52	28.40/10.23	28.05/0.88	29.22/15.31
Lena	26.93/0.73	24.59/0.03	29.47/13.22	30.58/8.90	30.58/0.81	31.37/15.19
Mandrill	21.40/0.64	20.59/0.06	22.72/13.58	21.78/9.57	22.58/0.72	23.30/15.58
Peppers	26.74/0.81	24.89/0.08	29.49/13.14	29.74/8.91	29.80/0.88	31.52/15.03
Mountain	19.23/0.95	17.69/0.09	20.11/15.24	18.45/11.12	20.03/1.05	20.42/17.47
Frog	23.23/0.94	22.35/0.16	24.00/16.07	24.40/13.37	24.69/1.09	24.69/21.14
Gold-hill	25.90/0.69	24.26/0.06	27.52/13.41	28.70/9.54	28.82/1.09	29.09/21.14
Average	24.12/0.81	22.55/0.08	25.88/14.03	26.01/10.23	26.31/0.89	27.09/16.19

reported in the last column of Table 1 include the running times of the BM3D denoising method.

4 Conclusion

In this paper, we derived the PLMMSE estimator and showed that it depends only on the joint second-order statistics of X and Y , rendering it applicable in a wide variety of situations. We demonstrated the utility of our approach in sparse signal recovery from a measurement pair. In the context of image enhancement from blurred/noisy image pairs, we showed that PLMMSE estimation performs close to state-of-the-art algorithms while running much faster.

References

1. Costa, O.L.V.: Linear minimum mean square error estimation for discrete-time Markovian jump linear systems. *IEEE Trans. Autom. Control* **39**(8) (1994) 1685–1689
2. Schniter, P., Potter, L.C., Ziniel, J.: Fast Bayesian matching pursuit. In: *Information Theory and Applications Workshop (ITA'08)*. (2008) 326–333
3. Soussen, C., Idier, J., Brie, D., Duan, J.: From bernoulli-gaussian deconvolution to sparse signal restoration. *IEEE Trans. Signal Process.* (99) (2010) 4572–4584
4. Girolami, M.: A Variational method for learning sparse and overcomplete representations. *Neural Computation* **13**(11) (2001) 2517–2532
5. Härdle, W., Liang, H.: Partially linear models. In: *Statistical Methods for Biostatistics and Related Fields*. (2007) 87–103
6. Michaeli, T., Sigalov, D., Eldar, Y.: Partially linear estimation with application to sparse signal recovery from measurement pairs. *IEEE Trans. Signal Process.* (2011) Submitted.
7. Yuan, L., Sun, J., Quan, L., Shum, H.Y.: Image deblurring with blurred/noisy image pairs. In: *ACM SIGGRAPH 2007 papers*, ACM (2007) 1–10
8. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8) (2007) 2080–2095
9. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image restoration by sparse 3d transform-domain collaborative filtering. In: *SPIE Electronic Imaging*. (2008)