Blind Minimax Estimators
Improving on Least-Squares Estimation

Zvika Ben-Haim
Yonina C. Eldar
Technion - Israel Institute of Technology

Summary

- Consider the linear regression model, $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, in which $\mathbf{x}$ is deterministic.
- *Blind minimax:* Estimate a region containing the unknown value $\mathbf{x}$, then apply a minimax estimator designed for this region.
- Obtained estimators strictly dominate the least-squares estimator (i.e., always achieve lower MSE).

Problem Setting

System of measurements: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$

- $\mathbf{x}$ unknown, deterministic parameter vector
- $\mathbf{w}$ Gaussian noise: zero mean, known covariance $\mathbf{C_w}$
- $\mathbf{H}$ known system model
- $\mathbf{y}$ observation vector

Objective:

Construct an estimator $\hat{\mathbf{x}}$ to estimate $\mathbf{x}$ from measurements $\mathbf{y}$, such that the mean-squared error (MSE) is minimized:

$$\min E\left\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right\}$$

Note that $\mathbf{x}$ is deterministic, so that the Bayesian approach (i.e., Wiener filtering) is not applicable here!

Previous Work

Least-squares estimator (Gauss, 1821)

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{y}$$

- Unbiased
- Achieves constant MSE, given by $\epsilon_0 = \text{Tr}((\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1})$.
- Achieves Cramer-Rao lower bound: best unbiased estimator
- **Outperformed by many biased estimators!**

Stein-type estimators

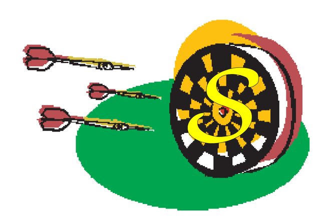
- Family of biased estimators outperforming the LSE
- James and Stein (1961) proposed an estimator for the i.i.d. case ($\mathbf{H} = \mathbf{I}, \mathbf{C_w} = \sigma^2\mathbf{I}$).
- Bock (1975) extended to general case:

$$\hat{\mathbf{x}}_{\text{Bock}} = \left(1 - \frac{\epsilon_0/\lambda_{\max} - 2}{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\hat{\mathbf{x}}_{\text{LS}}}\right)\hat{\mathbf{x}}_{\text{LS}}$$

where $\lambda_{\max}$ is the largest eigenvalue of $(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}$.

- Shrinkage estimator: inappropriate for some applications
- **Can other estimators achieve better improvement over LSE?**

Minimax Estimation



- Used when $\mathbf{x}$ is known to lie in a given parameter set $\mathcal{S}$
- Find linear estimator which minimizes the worst-case error within $\mathcal{S}$:

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{x}\in\mathcal{S}} E\left\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right\}$$

- Closed form solution is known for many types of parameter sets (Eldar et al., 2005)

Theorem: For any compact parameter set $\mathcal{S}$, a minimax estimator achieves lower MSE than the least-squares estimator, for all $\mathbf{x} \in \mathcal{S}$ (Ben-Haim and Eldar, 2005).

Blind Minimax Estimation

The idea:

Use minimax estimators even when a parameter set is unknown!

1. Estimate a parameter set $\mathcal{S}$ from the measurements
2. Apply a minimax estimator designed for $\mathcal{S}$

How does this work?

Blind minimax estimators work because a parameter set can be estimated far more accurately than the actual value of the parameter.

Spherical blind minimax estimator (SBME)

- Use spherical parameter set centered on the origin; estimate the radius from measurements
- Minimax estimator for a spherical set is given by

$$\hat{\mathbf{x}}_{\text{M}} = \frac{L^2}{L^2 + \epsilon_0}\hat{\mathbf{x}}_{\text{LS}}$$

where L is the radius.

- Note that $\|\mathbf{x}\|^2 = E\left\{\|\hat{\mathbf{x}}_{\text{LS}}\|^2\right\} - \epsilon_0$, so we choose to estimate $L^2 = \|\hat{\mathbf{x}}_{\text{LS}}\|^2 - \epsilon_0$, obtaining

$$\hat{\mathbf{x}}_{\text{SBM}} = \frac{\|\hat{\mathbf{x}}_{\text{LS}}\|^2 - \epsilon_0}{\|\hat{\mathbf{x}}_{\text{LS}}\|^2}\hat{\mathbf{x}}_{\text{LS}}$$

- This is a shrinkage estimator which reduces to Stein's estimator in the i.i.d. case.

Theorem: If the effective dimension $d > 4$, then the spherical blind minimax estimator achieves lower MSE than the least-squares estimator, *for all* values of $\mathbf{x}$.

- *Effective dimension:* the number of independent measurements, defined as $d = \text{Tr}(\mathbf{Q}^{-1})/\lambda_{\max}(\mathbf{Q}^{-1})$, where $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}$.

Ellipsoidal blind minimax estimator (EBME)

When the noise is colored, some measurements are more reliable than others. This information can be utilized using an ellipsoidal parameter set.

- Note that

$$\mathbf{Q}^{-1/2}\hat{\mathbf{x}}_{\text{LS}} = \mathbf{Q}^{-1/2}\mathbf{x} + \tilde{\mathbf{w}}$$

where $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}$ and $\tilde{\mathbf{w}}$ is white noise.

- Hence, rather than estimate $\|\mathbf{x}\|^2$ using $\|\hat{\mathbf{x}}_{\text{LS}}\|^2$, one can estimate $\mathbf{x}^*\mathbf{Q}^{-1}\mathbf{x}$ using $\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\text{LS}}$.

- This results in the ellipsoidal blind minimax estimator:

$$\hat{\mathbf{x}}_{\text{EBM}} = \mathbf{V}\,\text{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})\mathbf{V}^*\left(\mathbf{I} - \alpha\mathbf{Q}^{1/2}\right)\hat{\mathbf{x}}_{\text{LS}}$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m}\lambda_i^{-1/2}}{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\text{LS}} - k}$$

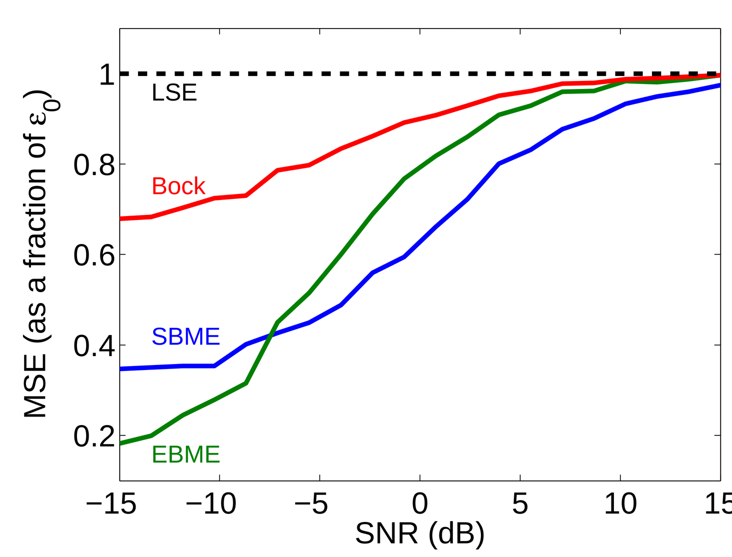
and k is the smallest integer $0 \le k \le m-1$ such that $\alpha < \lambda_{k+1}^{-1/2}$.

- The EBME is *not* a shrinkage estimator, which makes it appropriate for a wider range of applications.

Theorem: If $\text{Tr}(\mathbf{Q}^{-1/2})/\lambda_{\max}(\mathbf{Q}^{-1/2}) > 4$, then the ellipsoidal blind minimax estimator achieves lower MSE than the least-squares estimator, *for all* values of $\mathbf{x}$.

Numerical Results

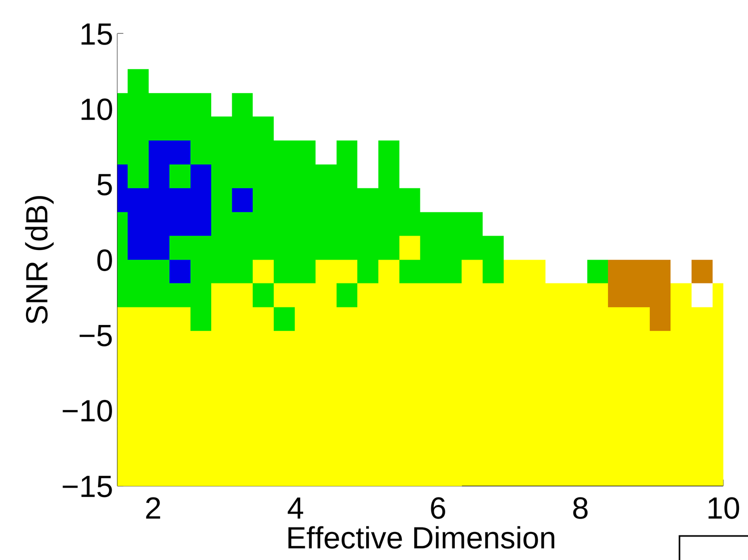
Estimator MSE comparison under typical operating conditions



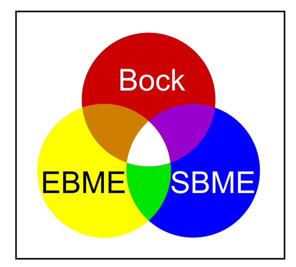
- 15 parameters
- 15 measurements
- Effective dimension 4.8
- x distributed normally i.i.d.

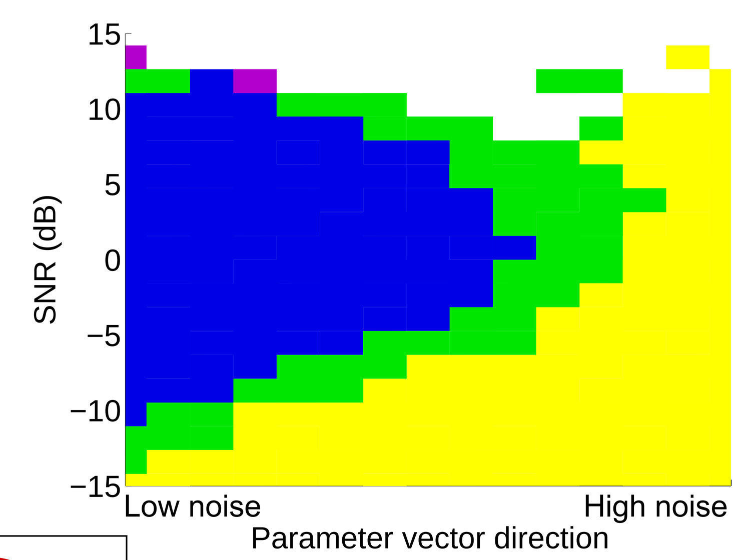
Optimal estimator under various operating conditions

Color indicates estimator achieving lowest MSE for given operating conditions. Combined colors indicate that two or more estimators achieve performance within 5% of optimal.



- 10 parameters
- 10 measurements
- x distributed i.i.d.





- 15 parameters
- 15 measurements
- Effective dimension 5

Discussion

- We provide two novel estimators based on the blind minimax technique.
- The proposed estimators achieve lower MSE than the least-squares estimator, under simple regularity conditions.
- The blind minimax estimators generally outperform Bock's estimator.
- Both estimators reduce to Stein's estimator in the i.i.d. case, although they are derived for a more general setting.

Where should these estimators be used?

- The SBME is a simple shrinkage estimator which can replace any usage of the least-squares estimator.
  Some examples are: linear prediction, system identification, and channel estimation.
- In some cases, a shrinkage estimator does not affect performance (e.g., a binary slicer, or an image reconstructor).
  This happens when the MSE is only an approximation of the true quality measure (bit error rate, subjective image quality, etc.).
  In such cases the EBME may provide improved performance.

References

Z. Ben-Haim and Y. C. Eldar, 2005, "Minimax estimators dominating the least-squares estimator," in *Proc. ICASSP 2005*, vol. IV, pp. 53-56.

M. E. Bock, 1975, "Minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 3, no. 1, pp. 209-218.

Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, 2005, "Robust mean-squared error estimation with bounded data uncertainties," *IEEE Trans. Signal Proc.*, vol. 53, no. 1, pp. 168-181.

K. F. Gauss, 1821, *Theory of Motion of Heavenly Bodies.*

W. James and C. Stein, 1961, "Estimation with quadratic loss," in *Proc 4th Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 311-319.

C. Stein, 1956, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 197-206.