Module – 6.2

STATISTICAL LEAST SQUARES

S. Lakshmivarahan

School of Computer Science

University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

STATISTICAL LEAST SQUARE ESTIMATE

- Given $z = Hx + v$, $z \in R^m$, $x \in R^n$
  - $E(v) = 0$
  - $E(vv^T) = R$ – known, S.P.D. => R^{-1} information matrix exists
  - x and v are NOT correlated
- Define residual $r(x) = z - Hx$
- Weight sum of squared residuals

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{1}{2}\,\mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) = \frac{1}{2}\,\|\mathbf{r}(\mathbf{x})\|^2_{\mathbf{R}^{-1}} \\
&= \frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{x})
\end{aligned}
$$

STATISTICAL LEAST SQUARE ESTIMATE (CONT'D)

- $\nabla f(x) = (H^T R^{-1} H)x - H^T R^{-1} z$ -> (1)
- $\nabla^2 f(x) = H^T R^{-1} H$ -> (2)
- $\hat{\mathbf{x}}_{LS} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}$ -> (3)

OBSERVATIONS

- Unbiasedness

$$\begin{aligned}\hat{\mathbf{x}}_{LS} &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} \\ &= \mathbf{x} + (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{v}\end{aligned}$$

and Z = Hx + V

$$\begin{aligned}E(\hat{\mathbf{x}}_{LS}) &= \mathbf{x} + (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}E(\mathbf{v}) \\ &= \mathbf{x}.\end{aligned}$$

- $\hat{x}_{LS}$ is unbiased
- Covariance of the estimate

$$\begin{aligned}COV(\hat{\mathbf{x}}_{LS}) &= E[(\hat{\mathbf{x}}_{LS} - \mathbf{x})(\hat{\mathbf{x}}_{LS} - \mathbf{x})^T] \\ &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}E(\mathbf{v}\mathbf{v}^T)\mathbf{R}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} \\ &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} \\ &= [\nabla^2 f(\mathbf{x})]^{-1}.\end{aligned}$$

OBSERVATIONS (CONT'D)

- Relation to Projection

$$
\begin{aligned}
\hat{\mathbf{z}} &= \mathbf{H}\hat{\mathbf{x}}_{LS} \\
&= \mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} \\
&= \mathbf{P}\mathbf{z}
\end{aligned}
$$

$$\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

  - Idempotent: $P^2 = P$
  - Non-symmetric: $P \neq P^T$
  - => P is an oblique projection
  - Note: when $R^{-1} = I$, $P = P^T$ is symmetric (orthogonal projection)

OBSERVATIONS (CONT'D)

- Uncorrelated noise: $\mathbf{R} = \sigma^2\mathbf{I}$

$$\hat{\mathbf{x}}_{LS} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{z}$$

$$COV(\hat{\mathbf{x}}_{LS}) = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$$

  - H^TH is symmetric
  - => $(H^TH)Q = Q\Lambda$ – Eigen decomposition of H^TH

    $(H^TH) = Q\Lambda Q^T$, $Q^TQ = QQ^T = \mathrm{I}$, $\Lambda = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$

    $(H^TH)^{-1} = Q\Lambda^{-1}Q^T$
  - => $\mathrm{tr}[COV(\hat{x}_{LS})] = \mathrm{tr}[\sigma^2(H^TH)^{-1}) = \sigma^2\mathrm{tr}[Q\Lambda^{-1}Q^T] = \sigma^2\mathrm{tr}[Q^TQ\Lambda^{-1}]$

    $= \sigma^2\mathrm{tr}[\Lambda^{-1}] = \sigma^2\sum_{i=1}^{n}\frac{1}{\lambda_i}$
  - If H^TH is nearly singular, then λ_i is close to 0. => $COV(\hat{x}_{LS})$ is large

OBSERVATIONS (CONT'D)

- Estimation of σ^2 : Let R = σ^2 I
  - Define the residue e (error in the estimate)

$$\begin{aligned} \mathbf{e} = \mathbf{z} - \hat{\mathbf{z}} &= \mathbf{z} - \mathbf{H}\hat{\mathbf{x}}_{LS} \\ &= (\mathbf{I} - \mathbf{P})\mathbf{z} \\ &= (\mathbf{I} - \mathbf{P})(\mathbf{H}\mathbf{x} + \mathbf{v}) \quad [(I - P)H = (H - PH) = H - H = 0] \\ &= (\mathbf{I} - \mathbf{P})\mathbf{v}, \end{aligned}$$

$$E(\mathbf{e}) = E[(\mathbf{I} - \mathbf{P})\mathbf{v}] = (\mathbf{I} - \mathbf{P})E(\mathbf{v}) = 0.$$

OBSERVATIONS (CONT'D)

- Estimation of σ^2 (cont'd)

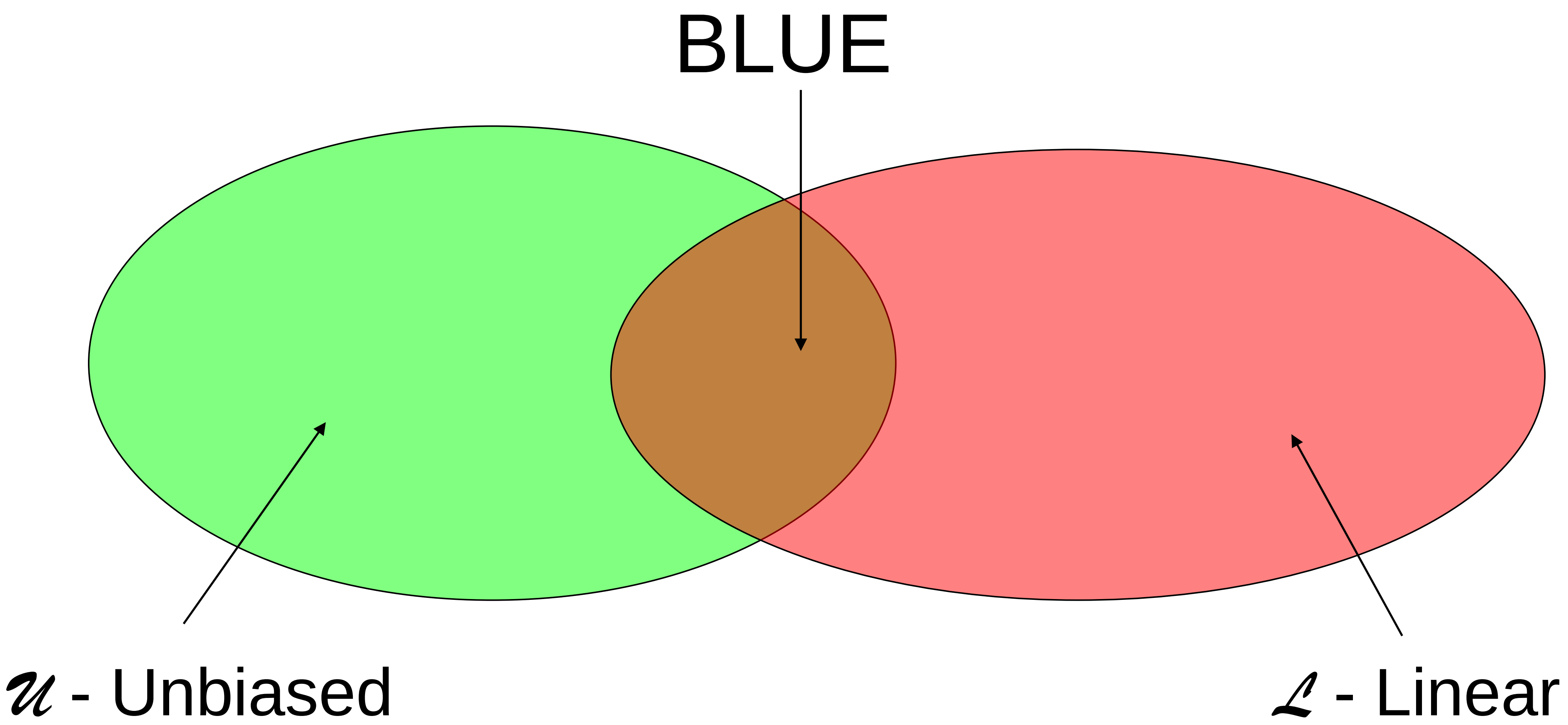
$$
\begin{aligned}
E(\mathbf{e}^T\mathbf{e}) &= E[\mathbf{v}^T(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})\mathbf{v}] \\
&= E[\mathbf{v}^T(\mathbf{I}-\mathbf{P})\mathbf{v}] && ((\mathbf{I}-\mathbf{P}) \text{ is idempotent}) \\
&= E[tr(\mathbf{v}^T(\mathbf{I}-\mathbf{P})\mathbf{v})] && (tr(a) = \text{a for scalar } a) \\
&= E[tr(\mathbf{v}\mathbf{v}^T(\mathbf{I}-\mathbf{P}))] && (tr(\mathbf{ABC}) = tr(\mathbf{CBA})) \\
&= \sigma^2 tr(\mathbf{I}-\mathbf{P}) \\
&= \sigma^2[tr(\mathbf{I}) - tr(\mathbf{P})] \\
&= \sigma^2(m-n)
\end{aligned}
$$

$P \in R^{mxm}$, $H^TH \in R^{nxn}$

$[tr[P] = tr[H(H^TH)^{-1}H^T] = tr[(H^TH)(H^TH)^{-1}] = tr[I]]$

- $\hat{\sigma^2} = \dfrac{\mathbf{e}^T\mathbf{e}}{(m-n)}$ is an unbiased estimate of σ^2

BLUE (BEST LINEAR UNBIASED ESTIMATE)



GAUSS-MARKOV THEOREM: OPTIMALITY OF LEAST SQUARES VERSION II

- Let x be the unknown being estimated, $X \in R^n$
- Pick μ and define $\Phi(x) = \mu^T x$, $\mu \in R^n$
- Consider the problem of estimating $\Phi(x)$
- We are seeking a linear **unbiased estimate** for $\Phi(x)$
- z is the data and let $a^T z$ be an estimator for $\Phi(x)$, $Z \in R^m$, $a \in R^m$

$$E[a^T z] = E[a^T(Hx+v)] = a^T H E(x) + a^T E(v) = a^T H x$$

GAUSS-MARKOV THEOREM (CONT'D)

- => 1) $a^T z$ is unbiased only if
  - $\Phi(x) = \mu^T x = E(a^T z) = a^T H x$
  - => $\mu^T = a^T H$ or $H^T a = \mu$
- 2) Since it is **unbiased**, M.S.E. = variance
  - $\text{var}(a^T z) = E[\, a^T z - E(a^T z)]^2$

$$= E[\, a^T (Hx+v) - a^T Hx]^2$$

$$= E[a^T v]^2$$

$$= a^T E(vv^T)a$$

$$= a^T R a$$

GAUSS-MARKOV THEOREM (CONT'D)

- Seek to minimize a^TRa when $H^Ta = \mu$

  $L(a, \lambda) = a^TRa - \lambda^T(H^Ta - \mu)$, Lagrangian, $\lambda \in R^n$

  $\nabla_a L(a, \lambda) = 2Ra - H\lambda = 0$

  $\nabla_\lambda L(a, \lambda) = H^Ta - \mu = 0$

  $\therefore a = ½(R^{-1}H\lambda)$, $H^Ta = \mu$

  $½(H^TR^{-1}H)\lambda = \mu$

  $\lambda = 2[H^TR^{-1}H]^{-1}\mu$

  $a = R^{-1}H(H^TR^{-1}H)^{-1}\mu$

GAUSS-MARKOV THEOREM (CONT'D)

- ∴ Linear, unbiased minimum variance estimate of $\Phi(x) = \mu^T x$ is

  $a^T z = \mu^T (H^T R^{-1} H)^{-1} H^T R^{-1} z = \mu^T \hat{X}_{LS}$

  Least squares estimate

- ∴ If $\mu = (1,0,\ldots.,0)^T$
- => is the best estimate of x_1 and so on.

NOTE

- If v is N(0,R), then $\hat{X}_{LS}$ is the best among all estimators - Rao-Blackwell Theorem.

- If v is not Gaussian, there exists non-linear estimates who's variance is smaller than linear estimate.

REFERENCES

- J. L. Melsa and D. L. Cohn (1978) Decision and Estimation Theory, *McGraw Hill*
- Also refer to chapter 14 in LLD (2006)