Improved Nonparametric Spectral Estimators

It follows from the bias-variance analysis of the periodogram that the values of the periodogram are asymptotically ($N \gg 1$) uncorrelated random variables whose means and standard deviations are both equal to the corresponding true power spectral density value.

That is,

$$E\left[\hat{\Gamma}_p(f)\right] \longrightarrow \Gamma(f)$$

$$\left(E\left[\left|\hat{\Gamma}_p(f) - E\left[\hat{\Gamma}_p(f)\right]\right|^2\right]\right)^{\frac{1}{2}} \longrightarrow \Gamma(f)$$

as $N \longrightarrow \infty$

Even as N increases, the values $\hat{\Gamma}_p(f)$ continue to fluctuate around the true value $\Gamma(f)$. Furthermore, because the values $\hat{\Gamma}_p(f)$ are asymptotically uncorrelated, the periodogram's errors will exhibit an erratic, noise-like behavior.

Recall that the overall MSE is decomposed into bias squared and variance. Asymptotically, the periodogram is doing well with the bias, but performs very poorly in terms of variance.

<u>Improved spectral estimators</u> decrease the variance at the expense of increasing the bias — the classic bias-variance trade-off in estimation theory.

The Blackman-Tukey Spectral Estimator

(Blackman and Tukey, 1959)

Reasons for Poor Performance of Periodogram

1. poor accuracy of $\hat{\gamma}(k)$ for large k $(k \sim N)$

2. large number of (small) covariance estimation errors that are summed up in forming $\hat{\Gamma}_p(f)$:

$$\hat{\Gamma}_p(f) = \sum_{k=-(N-1)}^{N-1} \hat{\gamma}(k) e^{-j2\pi f k}$$

Both of these effects can be reduced by truncating the sum in the definition of $\hat{\Gamma}_p(f)$. This leads to the Blackman-Tukey (BT) estimator:

$$\hat{\Gamma}_{BT}(f) = \sum_{k=-(M-1)}^{M-1} w(k)\, \hat{\gamma}(k)\, e^{-j2\pi f k}$$

where $M < N$ and $w(k)$ is a smoothly decaying window function.

Since $w(k)$ weights the "lags" of the covariance estimator, it is called a <u>lag window</u>.

<u>Properties of BT Lag Window:</u>

(i) $\{w(k)\}$ is an even function

$$w(k) = w(-k)$$

(ii) $w(0) = 1$

(iii) $w(k) = 0$ for $|k| \geq M$

(iv) $\{w(k)\}$ decays smoothly to zero with $|k|$ increasing

<u>Ex.</u> $\{w(k)\} \equiv$ rectangular window on $[-(M-1), \ldots, M-1]$

Let $W(f)$ denote the DTFT of $\{w(k)\}$:

$$W(f) = \sum_{k=-\infty}^{\infty} w(k)\, e^{-j2\pi fk}$$

$$= \sum_{k=-(M-1)}^{M-1} w(k)\, e^{-j2\pi fk}$$

Then

$$\hat{\Gamma}_{BT}(f) = \sum_{k=-\infty}^{\infty} w(k)\, \hat{r}(k)\, e^{-j2\pi fk}$$

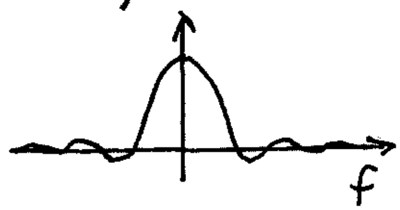
$$\overset{\text{DTFT}}{\longleftrightarrow} \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{\Gamma}_p(\nu)\, W(f-\nu)\, d\nu$$

That is,

$$\hat{\Gamma}_{BT}(f) = \hat{\Gamma}_p(f) * W(f)$$

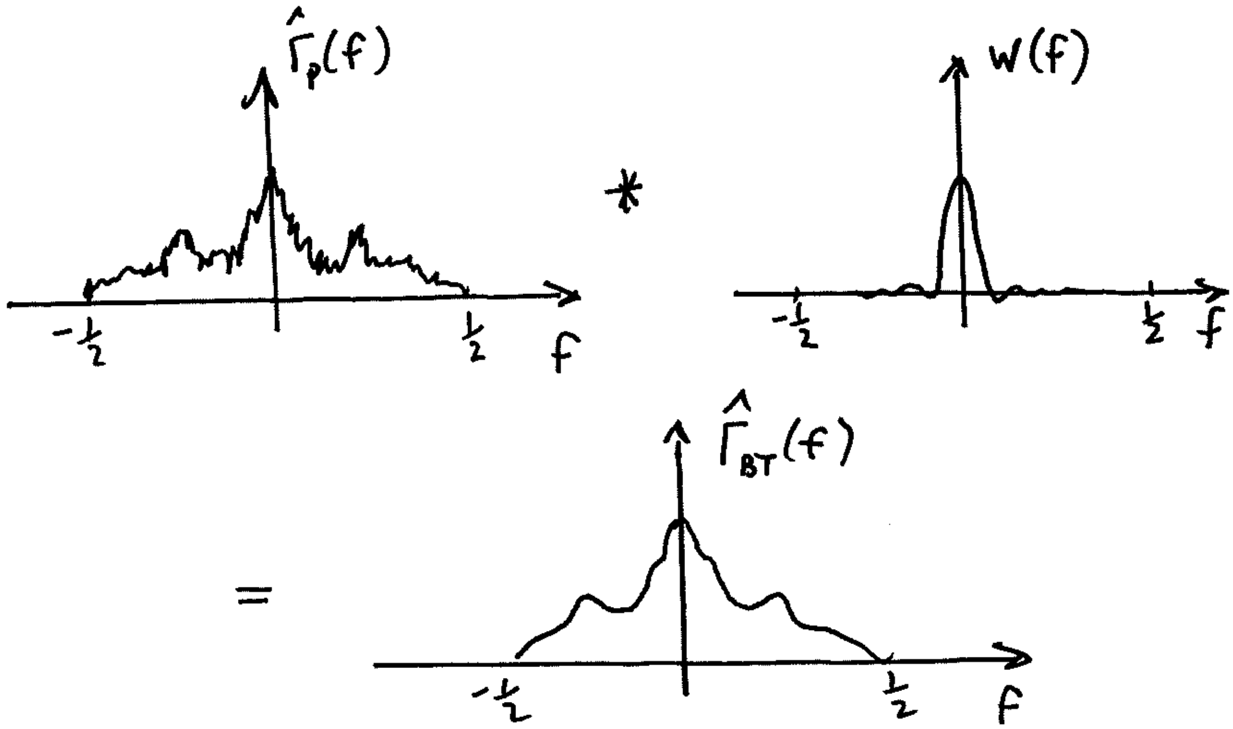
We can interpret the BT estimator as a "locally" weighted average of the periodogram.

$$\hat{\Gamma}_{BT}(f) = \hat{\Gamma}_p(f) * W(f)$$



fairly concentrated about $f=0$

Since $W(f)$ acts as a window (or weighting) in the frequency domain, its often called a <u>spectral window</u>.



Smoothing $\hat{\Gamma}_p(f)$ with the spectral window $W(f)$ reduces the "noisiness" (variance) of periodogram.

On the other hand, smoothing reduces the spectral resolution of the BT estimator. The resolution of $\hat{\Gamma}_{BT}$ is dictated by the width of the main lobe of $W(f)$.

Bias of BT Estimator

$$E\left[\hat{\Gamma}_{BT}(f)\right] = E\left[\hat{\Gamma}_{P}(f)\right] * W(f)$$

$$\longrightarrow \Gamma(f) * W(f)$$

as $N \longrightarrow \infty$

Thus, the BT estimator is asymptotically biased (unless $W(f) \longrightarrow \delta(f)$).

Recall that the effects of windowing are "smearing" or blurring due to the main lobe of $W(f)$ and spectral "leakage" due to the side lobes of $W(f)$. Also, recall that the width of the main lobe is $O(\frac{1}{M})$, for a length M window. So, the BT estimator has a spectral resolution of $O(\frac{1}{M})$, rather than $O(\frac{1}{N})$.

Variance of BT Estimator

Recall from our analysis of the periodogram variance (white-noise case, p. 65)

$$E\left[\hat{\Gamma}_p(f_1)\hat{\Gamma}_p(f_2)\right] = \sigma^4 + \sigma^4\left(\frac{\sin(\pi(f_1-f_2)N)}{N\sin(\pi(f_1-f_2))}\right)^2$$

and it follows from our analysis of the more general stationary Gaussian case (p. 67) that in general

$$E\left[\hat{\Gamma}_p(f_1)\hat{\Gamma}_p(f_2)\right] \to \Gamma(f_1)\Gamma(f_2) + \Gamma(f_1)\Gamma(f_2)\left(\frac{\sin(\pi(f_1-f_2)N)}{N\sin(\pi(f_1-f_2))}\right)^2$$

as $N \longrightarrow \infty$.

Consider

$$E\left[\left(\hat{\Gamma}_{BT}(f)\right)^2\right] = E\left[\int_{-\frac{1}{2}}^{\frac{1}{2}}\hat{\Gamma}_p(\nu_1)W(f-\nu_1)\,d\nu_1 \cdot \int_{-\frac{1}{2}}^{\frac{1}{2}}\hat{\Gamma}_p(\nu_2)W(f-\nu_2)\,d\nu_2\right]$$

$$E\left[\left(\hat{\Gamma}_{BT}(f)\right)^2\right]$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} E\left[\hat{\Gamma}_p(\nu_1)\,\hat{\Gamma}_p(\nu_2)\right] W(f-\nu_1)\,W(f-\nu_2)\,d\nu_1\,d\nu_2$$

$$\rightarrow \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma(\nu_1)\Gamma(\nu_2)\,W(f-\nu_1)\,W(f-\nu_2)\,d\nu_1\,d\nu_2$$

$$+ \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma(\nu_1)\Gamma(\nu_2) \left(\frac{\sin(\pi(\nu_1-\nu_2)N)}{N\sin(\pi(\nu_1-\nu_2))}\right)^2 W(f-\nu_1)\,W(f-\nu_2)\,d\nu_1\,d\nu_2$$

as $N \rightarrow \infty$

The first term above is just

$$\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma(\nu)\,W(f-\nu)\,d\nu\right)^2$$

i.e., the asymptotic mean of $\hat{\Gamma}_{BT}$ squared.

Thus, the second term is the variance term.

Now, for $N >> M$

$$\left(\frac{\sin(\pi f N)}{N \sin(\pi f)} \right)^2$$

is much more narrow (concentrated) than $W(f)$, and it "acts like" $\frac{1}{N}\delta(f)$. That is,

$$\left(\frac{\sin(\pi f N)}{N \sin(\pi f)} \right)^2 \approx \frac{1}{N}\, \delta(f)$$

Therefore, the variance is approximately

$$var\left(\hat{\Gamma}_{BT}(f) \right) \approx \frac{1}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma^2(\nu_1)\, W^2(f-\nu_1)\, d\nu_1$$

This may be further simplified if we assume that $\Gamma(f)$ is smoother than $W(f)$, in which case

$$var\left(\hat{\Gamma}_{BT}(f) \right) \approx \frac{\Gamma^2(f)}{N} \cdot \int_{-\frac{1}{2}}^{\frac{1}{2}} W^2(\ )\, d\nu$$

By Parseval's Theorem

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} W^2(\nu)\, d\nu = \sum_{k=-(M-1)}^{M-1} w^2(k)$$

$$\sim O(M)$$

and we find

$$\operatorname{var}\left(\hat{\Gamma}_{BT}(f)\right) \approx \frac{M}{N}\, \Gamma^2(f) \cdot \text{const.}$$

for $N \gg M$

depends on window, but is $O(1)$

<u>Conclusion</u>:

BT estimator reduces spectral resolution to $\frac{1}{M}$ (increased bias)

BT estimator reduces variance by a factor of $\frac{M}{N}$

Window Design Considerations

Nonnegativity:

To guarantee that $\hat{\Gamma}_{BT} \geq 0$
we require that $W \geq 0$.

Time-Bandwidth Product:

effective time width:

$$N_e \equiv \frac{\sum_{k=-(M-1)}^{M-1} w(k)}{w(0)}$$

effective bandwidth:

$$\beta_e \equiv \frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} W(f)\, df}{W(0)}$$

The effective time-bandwidth product:

$$N_e \, \beta_e = 1$$

Also,

$$N_e \sim O(M)$$

$$\beta_e \sim O\left(\frac{1}{M}\right)$$

$\Rightarrow$ spectral resolution of $W(f)$ is $O\left(\frac{1}{M}\right)$

$\Rightarrow$ variance reduction is $O\left(\frac{M}{N}\right)$

Trade-off between resolution and variance:

$M \uparrow \Rightarrow$ resolution (good), variance (bad) $\uparrow$

$M \downarrow \Rightarrow$ resolution (bad), variance (good) $\downarrow$

Window Shape

We must have $W(f) \geq 0$.

Ex.

Rectangular window ?

Triangular window ?
(Bartlett)

Common Windows

The windows satisfy $w(k) \equiv 0$ for $|k| \geq M$, and $w(k) = w(-k)$; the defining equations below are valid for $0 \leq k \leq (M-1)$.

Window Name	Defining Equation	Approx. Main Lobe Width (radians)	Sidelobe Level (dB)
Rectangular	$w(k) = 1$	$2\pi/M$	-13
Bartlett	$w(k) = (M-k)/M$	$4\pi/M$	-25
Hanning	$w(k) = .5 + .5\cos(\pi k/M)$	$4\pi/M$	-31
Hamming	$w(k) = .54 + .46\cos(\pi k/(M-1))$	$4\pi/M$	-41
Blackman	$w(k) = .42 + .5\cos(\pi k/(M-1))$ $+.08\cos(2\pi k/(M-1))$	$6\pi/M$	-57

From "Introduction to Spectral Analysis"
Stoica & Moses, Prentice Hall, 1997

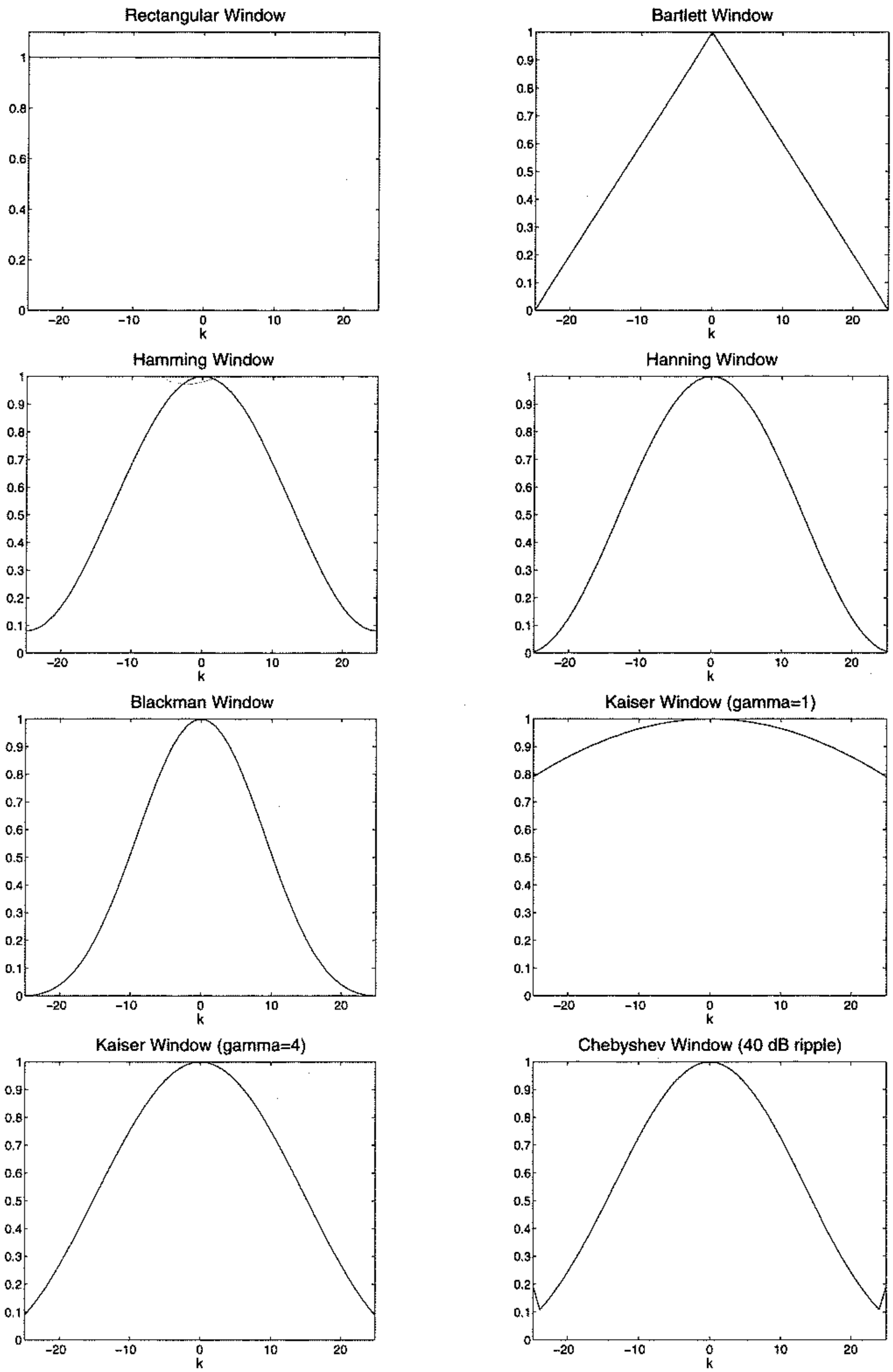


Figure 2.3. Some common window functions (shown for $M = 26$). The Kaiser window uses $\gamma = 1$ and $\gamma = 4$ and the Chebyshev window is designed for a -40 dB sidelobe level.

from Stoica & Moses

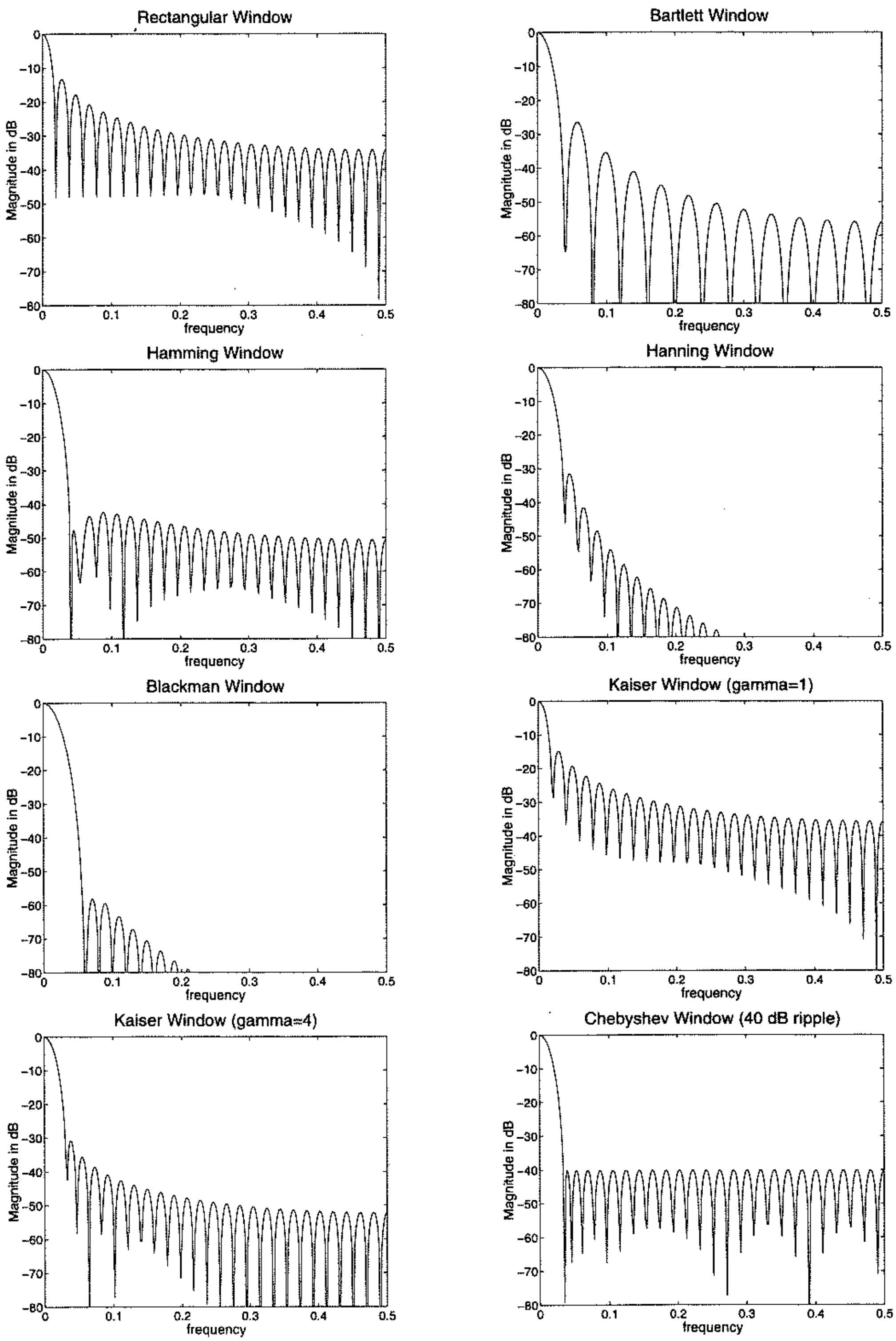


Figure 2.4. The DTFTs of the window functions in Figure 2.3.

90

Periodogram Averaging Methods

Basic Idea : (Bartlett Method)

1. Split data $x(n)$, $n = 0, \ldots, N-1$ into M segments of length K, $N = K \cdot M$.

2. Compute periodogram of each segment.

3. Average periodogram estimates. (variance reduced by M/N)

(Also see Welch method - p. 911 Proakis & Manolakis)

Averaging methods can be interpreted as special cases of the BT estimator. However, they generally perform slightly worse than the BT estimator, equipped with a triangular window, in terms of the bias-variance trade-off.

Summary:

BT and Periodogram estimators improve on the performance of the basic periodogram by reducing the variance by a factor of $\frac{M}{N}$, at the expense of reducing the spectral resolution to $\frac{1}{M}$.

Classic bias-variance trade-off.

<u>Next:</u> Parametric spectral estimation