	man Prediction Step	Notation:
Recall prediction	step of sequential Bayesian estimation	$oldsymbol{x} = oldsymbol{x}_n$
	$p(\boldsymbol{x} \boldsymbol{y}_{-}) = \int p(\boldsymbol{x} \boldsymbol{x}_{-}) \frac{p(\boldsymbol{x}_{-} \boldsymbol{y}_{-})}{p(\boldsymbol{x}_{-} \boldsymbol{y}_{-})} d\boldsymbol{x}_{-}$	$egin{aligned} oldsymbol{y} &= oldsymbol{y}_n \ oldsymbol{u} &= oldsymbol{u}_n \end{aligned}$
	$ \frac{p(\boldsymbol{x} \boldsymbol{y}_{-})}{\text{Predicted}} = \int p(\boldsymbol{x} \boldsymbol{x}_{-}) \frac{p(\boldsymbol{x}_{-} \boldsymbol{y}_{-})}{p(\boldsymbol{x}_{-} \boldsymbol{y}_{-})} d\boldsymbol{x}_{-} $ Previous posterior pdf	$oldsymbol{x} = oldsymbol{x}_{n-1}$
	$= \int p(\boldsymbol{x} \boldsymbol{x}_{-},\boldsymbol{y}_{-})p(\boldsymbol{x}_{-} \boldsymbol{y}_{-})d\boldsymbol{x}_{-} \qquad \text{(Markov Assumptions)}$	$egin{aligned} oldsymbol{y} &= oldsymbol{y}_{1:n-} \ oldsymbol{G} &= oldsymbol{G}_n \end{aligned}$
	$=\int p(oldsymbol{x},oldsymbol{x}_{-} oldsymbol{y}_{-})\mathrm{d}oldsymbol{x}_{-}$	
Note that $p(x y$	$(oldsymbol{y}) = \mathcal{N}(oldsymbol{\mu_{x y}}, oldsymbol{\Sigma_{xx y}})$	
Part I: First, we sh	now that, given $m{y}$ , $m{w} = [m{x}^{ m T}, m{x}^{ m T}]^{ m T}$ follows a joint Gaussian distribution, i.e.,	
	$p(\boldsymbol{x}, \boldsymbol{x} \boldsymbol{y}) = p(\boldsymbol{w} \boldsymbol{y}) \propto \mathcal{N}(\boldsymbol{\mu_w}, \boldsymbol{\varSigma_{ww}})$	
with mean and co	ovariance given by	
$\mu_w = egin{bmatrix} \mu_{x y } \ \mu_{x y } \end{bmatrix}$	$egin{aligned} oxed{iggreen} oxed{iggreen} oxed{iggreen} oxed{iggreen} oxed{iggreen} oxed{oxed} oxed{egin{aligned} oxed{oxed} oxed{egin{aligned} oxed{oxed} oxed{oxed} oxed{oxed} oxed{oxed} oxen{oxed} oxed{oxed} oxed{oxed} oxed{oxed} oxed{oxed} oxen{oxed} oxed{oxed} oxen{oxed} oxed{oxed} oxen{oxed} oxen{oxen} oxen{oxed} oxen{oxed} oxen{oxed} oxen{oxed} oxen{oxen} ox oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{oxen} oxen{ox oxen} oxen{oxen} ox ox{oxen} ox ox{oxen} ox ox{ox{} ox ox{} ox ox ox{} ox ox ox{} ox ox{} ox ox{} ox ox ox{} ox ox{} ox ox{}$	$egin{aligned} oldsymbol{\Sigma_{xx y}} oldsymbol{G}^{ ext{T}} \ oldsymbol{\Sigma_{xx y}} oldsymbol{G}^{ ext{T}} + oldsymbol{\Sigma_{u}} \end{aligned}$

1

Part I (cnt.): Note that  $p(x|x_-,y_-)=p(x|x_-)=\mathcal{N}(Gx_-,\Sigma_u)$  and  $p(x_-|y_-)=\mathcal{N}(\mu_{x_-|y_-},\Sigma_{x_-x_-|y_-})$ are both Gaussian distributed. Thus, we have that  $p(w|y_-) = p(x, x_-|y_-) = p(x|x_-, y_-)p(x_-|y_-)$  follows a joint Gaussian distribution (as derived in a previous class; see slides 5). The missing parameters of the joint Gaussian distribution can be calculated as follows  $\mathbf{\Sigma}_{oldsymbol{x}oldsymbol{x}_{-}|oldsymbol{y}_{-}} = \mathrm{E}ig[(oldsymbol{x} - oldsymbol{\mu}_{oldsymbol{x}|oldsymbol{y}_{-}})(oldsymbol{x}_{-} - oldsymbol{\mu}_{oldsymbol{x}_{-}|oldsymbol{y}_{-}})^{\mathrm{T}}ig]$  $\mu_{m{x}|m{y}_-} = \mathrm{E}(m{G}m{x}_- + m{u}) = m{G}m{\mu}_{m{x}_-|m{y}_-}$  $= G \operatorname{E}[x_{-}x_{-}^{\operatorname{T}}] - G \operatorname{E}[x_{-}] \mu_{x_{-}|y_{-}}^{\operatorname{T}}$  $=G(oldsymbol{\Sigma_{x_-x_-|y_-}}+oldsymbol{\mu_{x_-|y_-}}oldsymbol{\mu_{x_-|y_-}}^{ ext{T}})$  $oldsymbol{\Sigma_{xx|y_-}} = \mathrm{E}ig[(x-oldsymbol{\mu_{x|y_-}})(x-oldsymbol{\mu_{x|y_-}})^{\mathrm{T}}ig]$  $-\,G\mu_{oldsymbol{x}_-|oldsymbol{y}_-}\mu_{oldsymbol{x}_-|oldsymbol{y}_-}^{\mathrm{T}}$  $= \mathrm{E}ig[oldsymbol{x} oldsymbol{x}^{\mathrm{T}}ig] - oldsymbol{\mu}_{oldsymbol{x}|oldsymbol{y}_{-}} oldsymbol{\mu}_{oldsymbol{x}|oldsymbol{y}_{-}}^{\mathrm{T}}$  $=\mathbb{E}ig[(oldsymbol{G}oldsymbol{x}_- + oldsymbol{u})(oldsymbol{G}oldsymbol{x}_- + oldsymbol{u})^{ ext{T}}ig] - oldsymbol{G}oldsymbol{\mu}_{oldsymbol{x}_- | oldsymbol{y}_- oldsymbol{\mu}_{oldsymbol{x}_- | oldsymbol{y}_- oldsymbol{G}^{ ext{T}}}ig]$  $oldsymbol{\Sigma_{x_-x|y_-}} = oldsymbol{\Sigma_{xx_-|y_-}}^{ ext{T}}$  $= oldsymbol{G} \mathrm{E}[oldsymbol{x}_{-}oldsymbol{x}_{-}^{\mathrm{T}}] oldsymbol{G}^{\mathrm{T}} + \mathrm{E}ig[oldsymbol{u}oldsymbol{u}^{\mathrm{T}}ig] - oldsymbol{G}oldsymbol{\mu}_{oldsymbol{x}_{-}|oldsymbol{y}_{-}}oldsymbol{\mu}_{oldsymbol{x}_{-}|oldsymbol{y}_{-}}oldsymbol{G}^{\mathrm{T}}$  $= G \Sigma_{oldsymbol{x}_- oldsymbol{x}_-} |_{oldsymbol{y}_-} G^{ ext{T}} + \Sigma_{oldsymbol{u}oldsymbol{u}}$  $\mathrm{E}[oldsymbol{x}_{-}oldsymbol{x}^{\mathrm{T}}] = oldsymbol{\Sigma}_{oldsymbol{x}-oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{y}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-oldsymbol{\mu}}^{\mathrm{T}}_{oldsymbol{x}-olds$ Note that here  $\ \mathrm{E}[\cdot] = \int \int \cdot p(m{x}_-, m{u} | m{y}_-) \, \mathrm{d} m{x} \, \mathrm{d} m{u}$  $= \int \int \cdot p(\boldsymbol{x}_{-}|\boldsymbol{y}_{-}) p(\boldsymbol{u}) d\boldsymbol{x} d\boldsymbol{u}$ (Markov Assumptions)

2

Part II: Next, we get the mean  $\mu_{w|y_-}$  and covariance  $\Sigma_{xw|y_-}$  of the marginal Gaussian  $p(x|y_-) = \mathcal{N}(\mu_{w|y_-}, \Sigma_{xw|y_-})$  from the joint Gaussian  $p(x, x_-|y_-) = p(w|y_-) \propto \mathcal{N}(\mu_w, \Sigma_{ww})$ . These parameters can be directly extracted from the mean  $\mu_w$  and the covariance  $\Sigma_{ww}$  of the joint Gaussian (as derived in a previous class; see slides 5).

In particular, in this way we obtain  $\Sigma_{xw|y_-} = G\Sigma_{x_-x_-|y_-}G^T + \Sigma_{uu}$ 

3

Derivation of Kalman Update Step			Notation:
Recall update step of sequential Bayesian estimation			$oldsymbol{x} = oldsymbol{x}_n$
			$oldsymbol{y} = oldsymbol{y}_n$
$p(oldsymbol{x} oldsymbol{y},oldsymbol{y})$	$\propto p(oldsymbol{y}   oldsymbol{x}) \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$		$oldsymbol{v} = oldsymbol{v}_n$
Posterior pdf			$oldsymbol{x} = oldsymbol{x}_{n-1}$
	$=p(oldsymbol{y} oldsymbol{x},oldsymbol{y})p(oldsymbol{x} oldsymbol{y})$	(Markov Assumptions)	$oldsymbol{y} = oldsymbol{y}_{1:n}$ .
		(Warkov Assumptions)	$oldsymbol{H} = oldsymbol{H}_n$
	$=p(\boldsymbol{x},\boldsymbol{y} \boldsymbol{y}_{-})$		
Note that $p({m x}   {m y}) = \mathcal{N}({m \mu}_{m x}, {m \Sigma}_{{m x}{m x}})$			
<b>Part I:</b> First, we show that, given $y$ , $z$	$= [oldsymbol{x}^{ ext{T}}, oldsymbol{y}^{ ext{T}}]^{ ext{T}}$ follows a	joint Gaussian distribution, i.e.,	
	$p(oldsymbol{x}, oldsymbol{y}   oldsymbol{y}) = p(oldsymbol{z}   oldsymbol{y}$	$(oldsymbol{\mu}_{oldsymbol{z}}) = \mathcal{N}(oldsymbol{\mu}_{oldsymbol{z}}, oldsymbol{\Sigma}_{oldsymbol{z},oldsymbol{z}})$	
	1 ( ) 0   0 ) 1 (   0	/ 22/	
with mean and covariance given by			
F 7 F	7		
$\mu_x =  \mu_x  =  \mu $	$x$ $\sum_{x} =$	$egin{array}{c cccc} oldsymbol{\Sigma_{xx}} & oldsymbol{\Sigma_{xy}} & oldsymbol{\Sigma_{xy}} & oldsymbol{\Sigma_{xx}} & oldsymbol{H} oldsymbol{\Sigma_{xx}} & oldsymbol{X_{xx}} & oldsymbol{H} oldsymbol{\Sigma_{xx}} & oldsymbol{H} oldsymbol{\Sigma_{xx}} & oldsymbol{H} oldsymbol{X_{xx}} & oldsymbol{H} oldsymbol{X_{xx}} & $	$oldsymbol{arSigma_{xx}} oldsymbol{H}^{ ext{T}}$
$\mu_y$ $H$	$\mu_x$	$egin{bmatrix} oldsymbol{\Sigma}_{yx} & oldsymbol{\Sigma}_{yy} \end{bmatrix} & egin{bmatrix} H oldsymbol{\Sigma}_{xx} & H \end{bmatrix}$	$oldsymbol{arSigma_{xx}H^{\mathrm{T}}} + oldsymbol{arSigma_{vv}}$

1

Part I (cnt.): Note that  $p(y|x,y_-) = p(y|x) = \mathcal{N}(Hx,\Sigma_v)$  and  $p(x|y_-) = \mathcal{N}(\mu_x,\Sigma_{xx})$  are both Gaussian distributed. Thus, we have that  $p(z|y_-) = p(x,y|y_-) = p(y|x,y_-)p(x|y_-)$  follows a joint Gaussian distribution (as derived in a previous class; see slides 5). The missing parameters of the joint Gaussian distribution can be calculated as follows

$$egin{aligned} egin{aligned} egi$$

Note that here 
$$E[\cdot] = \int \int \cdot p(\boldsymbol{x}, \boldsymbol{v} | \boldsymbol{y}_-) d\boldsymbol{x} d\boldsymbol{v}$$
 
$$= \int \int \cdot p(\boldsymbol{x} | \boldsymbol{y}_-) p(\boldsymbol{v}) d\boldsymbol{x} d\boldsymbol{v} \qquad \text{(Markov Assumptions)}$$

5

Part II: Next, we calculate the mean  $\mu_{x|y}$  and covariance  $\Sigma_{xx|y}$  of  $p(x|y,y_-) = \mathcal{N}(\mu_{x|y},\Sigma_{xx|y})$  from  $p(x,y|y_-) = p(z|y_-) \propto \mathcal{N}(\mu_z,\Sigma_{zz})$  by using the Schur complement (as derived in a previous class; see slides 5). In this way, we obtain

$$egin{aligned} \mu_{x|y} &= \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y) \ \Sigma_{xx|y} &= \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \end{aligned}$$

Equivalently, we can write

$$egin{aligned} \mu_{x|y} &= \mu_x + K(y - \mu_y) \ & \Sigma_{xx|y} &= \Sigma_{xx} - KH\Sigma_{xx} \end{aligned}$$

where we have introduced the Kalman gain

$$egin{aligned} K &= oldsymbol{\Sigma_{xy}} oldsymbol{\Sigma_{yy}}^{-1} \ &= oldsymbol{\Sigma_{xx}} oldsymbol{H}^{ ext{T}} (oldsymbol{H} oldsymbol{\Sigma_{xx}} oldsymbol{H}^{ ext{T}} + oldsymbol{\Sigma_{v}})^{-1} \end{aligned}$$

6