5.27 We have:

$$\mathbf{s}(\mathbf{x},\boldsymbol{\xi}) = \mathbf{x} + \boldsymbol{\xi}$$

and

$$\boldsymbol{\tau} = \frac{\partial \mathbf{s}(\mathbf{x},\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \frac{\partial(\mathbf{x}+\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \mathbf{I}$$

and

$$\boldsymbol{\tau}' = \mathbf{0}$$

Obtaining the Second order Taylor approximation of $y(\mathbf{s}(\mathbf{x},\boldsymbol{\xi}))$:

$$y(\mathbf{s}(\mathbf{x},\boldsymbol{\xi})) = y(\mathbf{x}+\boldsymbol{\xi}) = y(\mathbf{x}) + \boldsymbol{\xi}^T \triangledown y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi}$$

Substituting into the mean error function (5.130) and expanding, we then have:

$$\tilde{E} = \frac{1}{2}\int\int\int \left\{ y(\mathbf{x}) + \boldsymbol{\xi}^T \triangledown y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi} - t \right\}^2 p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2\ p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$
$$+\frac{1}{2}\int\int\int \left\{ \boldsymbol{\xi}^T \triangledown y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi} \right\}^2 p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$
$$+\int\int\int \{y(\mathbf{x}) - t\} \left\{ \boldsymbol{\xi}^T \triangledown y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi} \right\} p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2\ p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$
$$+\frac{1}{2}\int\int\int \left\{ \left(\boldsymbol{\xi}^T \triangledown y(\mathbf{x})\right)^2 + \left(\frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi}\right)^2 + \boldsymbol{\xi}^T \triangledown y(\mathbf{x})\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi} \right\} p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$
$$+\int\int\int \{y(\mathbf{x}) - t\} \left\{ \boldsymbol{\xi}^T \triangledown y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \triangledown\triangledown y(\mathbf{x})\boldsymbol{\xi} \right\} p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

Ignoring higher order terms w.r.t $\boldsymbol{\xi}$, this becomes:

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\frac{1}{2}\int\int\int \left\{\boldsymbol{\xi}^T \nabla y(\mathbf{x})\right\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\int\int\int \{y(\mathbf{x}) - t\} \left\{\boldsymbol{\xi}^T \nabla y(\mathbf{x}) + \frac{1}{2}\boldsymbol{\xi}^T \nabla\nabla y(\mathbf{x})\boldsymbol{\xi}\right\} \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

Rearranging terms, we get:

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\int\int\int \{y(\mathbf{x}) - t\}\, \boldsymbol{\xi}^T \nabla y(\mathbf{x})\, p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\frac{1}{2}\int\int\int \left(\left\{\boldsymbol{\xi}^T \nabla y(\mathbf{x})\right\}^2 + \{y(\mathbf{x}) - t\}\, \boldsymbol{\xi}^T \nabla\nabla y(\mathbf{x})\boldsymbol{\xi}\right) \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

The second term goes to 0, as the mean of $\boldsymbol{\xi}$ is $\mathbf{0}$. So the expression becomes:

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\frac{1}{2}\int\int\int \left(\left\{\boldsymbol{\xi}^T \nabla y(\mathbf{x})\right\}^2 + \{y(\mathbf{x}) - t\}\, \boldsymbol{\xi}^T \nabla\nabla y(\mathbf{x})\boldsymbol{\xi}\right) \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

Calculating the integral w.r.t t:

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\frac{1}{2}\int\int \left(\left\{\boldsymbol{\xi}^T \nabla y(\mathbf{x})\right\}^2 + \{y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}\, \boldsymbol{\xi}^T \nabla\nabla y(\mathbf{x})\boldsymbol{\xi}\right) \; p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, d\boldsymbol{\xi}$$

As explained in the book, $y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]$ is a second order term w.r.t $\boldsymbol{\xi}$, which makes $\{y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}\, \boldsymbol{\xi}^T \nabla\nabla y(\mathbf{x})\boldsymbol{\xi}$ a higher order term that we can ignore, giving us:

$$= \frac{1}{2}\int\int\int \{y(\mathbf{x}) - t\}^2 \; p(t|\mathbf{x})\, p(\mathbf{x})\, p(\boldsymbol{\xi})\, d\mathbf{x}\, dt\, d\boldsymbol{\xi}$$

$$+\frac{1}{2}\int\int\left\{\boldsymbol{\xi}^T\nabla y(\mathbf{x})\right\}^2 p(\mathbf{x})\,p(\boldsymbol{\xi})\,d\mathbf{x}\,d\boldsymbol{\xi}$$

The second term is the regularizer, which can be further simplified as:

$$\begin{aligned}
&\frac{1}{2}\int\int\left\{\boldsymbol{\xi}^T\nabla y(\mathbf{x})\right\}^2 p(\mathbf{x})\,p(\boldsymbol{\xi})\,d\mathbf{x}\,d\boldsymbol{\xi}\\
&=\frac{1}{2}\int\int\nabla y(\mathbf{x})^T\boldsymbol{\xi}\boldsymbol{\xi}^T\nabla y(\mathbf{x})\,p(\mathbf{x})\,p(\boldsymbol{\xi})\,d\mathbf{x}\,d\boldsymbol{\xi}\\
&=\frac{1}{2}\int\nabla y(\mathbf{x})^T\left(\int\boldsymbol{\xi}\boldsymbol{\xi}^T\,p(\boldsymbol{\xi})\,d\boldsymbol{\xi}\right)\nabla y(\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x}
\end{aligned}$$

As per this definition, as $\mathbb{E}[\boldsymbol{\xi}]=\mathbf{0}$:

$$\int\boldsymbol{\xi}\boldsymbol{\xi}^T\,p(\boldsymbol{\xi})\,d\boldsymbol{\xi}=\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T]=cov[\boldsymbol{\xi}]=\mathbf{I}$$

$$\Longrightarrow\Omega=\frac{1}{2}\int\nabla y(\mathbf{x})^T\mathbf{I}\nabla y(\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x}$$

$$\Longrightarrow\Omega=\frac{1}{2}\int||\nabla y(\mathbf{x})||^2\,p(\mathbf{x})\,d\mathbf{x}$$

which is the same as the result in 5.135 that we wanted.