SELU integrals

Here we compute the integrals for the first and second moment moment of the next layer.

In[2]:=
```
Simplify[Integrate[
   λ (α * Exp[z] - α) * PDF[NormalDistribution[μ * ω, Sqrt[ν * τ]], z], {z, -Infinity, 0}] +
  Integrate[λ * z * PDF[NormalDistribution[μ * ω, Sqrt[ν * τ]], z], {z, 0, Infinity}],
 Assumptions → ν ∈ Reals && ν > 0 && τ ∈ Reals && τ > 0]
```

Out[2]=
$$\frac{1}{2}\lambda\left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}+\mu\omega+\mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-\alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

In[3]:=
```
Simplify[
 Integrate[(λ (α * Exp[z] - α)) ^2 * PDF[NormalDistribution[μ * ω, Sqrt[ν * τ]], z],
   {z, -Infinity, 0}] + Integrate[
   (λ * z) ^2 * PDF[NormalDistribution[μ * ω, Sqrt[ν * τ]], z], {z, 0, Infinity}],
 Assumptions → ν ∈ Reals && ν > 0 && τ ∈ Reals && τ > 0]
```

Out[3]=
$$\frac{1}{2}\lambda^2\left(\nu\tau+e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\mu\sqrt{\nu\tau}\,\omega+\mu^2\omega^2+\left(\nu\tau+\mu^2\omega^2\right)\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+\alpha^2\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-2e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{2\nu\tau+2\mu\omega}\alpha^2\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

Mean, second moment, variance

We define the mean, second moment and the variance in the next layer.

In[4]:=

firstMoment[μ_, ω_, ν_, τ_, λ_, α_] :=
$$\frac{1}{2}\lambda\left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}+\mu\omega+\mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-\alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

secondMoment[μ_, ω_, ν_, τ_, λ_, α_] :=
$$\frac{1}{2}\lambda^2\left(\nu\tau+e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\mu\sqrt{\nu\tau}\,\omega+\mu^2\omega^2+\left(\nu\tau+\mu^2\omega^2\right)\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+\alpha^2\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-2e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{2\nu\tau+2\mu\omega}\alpha^2\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

In[6]:=
```
variance[μ_, ω_, ν_, τ_, λ_, α_] :=
 secondMoment[μ, ω, ν, τ, λ, α] - firstMoment[μ, ω, ν, τ, λ, α] ^2
```

In[7]:=

Solve fixed point equations for alpha and lambda

Here, we solve the fixed point equations with respect to alpha and lambda.

In[8]:=
```
Solve[firstMoment[0, 0, 1, 1, λ, α] == 0, α]
```

Out[8]=
$$\left\{\left\{\alpha \to -\frac{\sqrt{\frac{2}{\pi}}}{-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]}\right\}\right\}$$

In[9]:=
```
α01 := -Sqrt[2 / Pi] / (Exp[1 / 2] * Erfc[1 / Sqrt[2]] - 1)
N[α01]
```

Out[10]= 1.67326

In[11]:=
```
Solve[variance[0, 0, 1, 1, λ, α01] == 1, λ]
```

Out[11]=
$$\left\{\left\{\lambda \to -\sqrt{\left(2\Big/\left(1-\frac{1}{\pi}+\frac{1}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{2\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{e\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]^2}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{2}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)}+\frac{2\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)}+\frac{2e^2\ \mathrm{Erfc}\left[\sqrt{2}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}\right)\right)}\right\},$$
$$\left\{\lambda \to \sqrt{\left(2\Big/\left(1-\frac{1}{\pi}+\frac{1}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{2\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{e\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]^2}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}-\frac{2}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)}+\frac{2\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)}+\frac{2e^2\ \mathrm{Erfc}\left[\sqrt{2}\right]}{\pi\left(-1+\sqrt{e}\ \mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)^2}\right)\right)}\right\}\right\}$$

In[12]:=
```
λ01 :=
 ((1 - Sqrt[Exp[1]] * Erfc[1 / Sqrt[2]]) * Sqrt[2 * Pi / (Exp[1] * Pi * Erfc[1 / Sqrt[2]] ^ 2 -
      2 * Sqrt[Exp[1]] (2 + Pi) Erfc[1 / Sqrt[2]] + 2 * Exp[1] ^ 2 Erfc[Sqrt[2]] + Pi + 2)])
```

In[13]:= `Simplify[λ01]`

Out[13]=
$$\left(1-\sqrt{e}\,\mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]\right)\sqrt{(2\pi)\Big/\left(2+\pi-2\sqrt{e}\,(2+\pi)\,\mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]+e\,\pi\,\mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\right]^2+2\,e^2\,\mathrm{Erfc}\left[\sqrt{2}\right]\right)}$$

In[14]:= `N[λ01]`

Out[14]= 1.0507

Fixed point for normalized weights

This is just a check whether the obtained parameters λ01 and α01 lead to the fixed point (0,1) for normalized weights.

In[15]:= `Simplify[firstMoment[0, 0, 1, 1, λ01, α01]]`

Out[15]= 0

In[16]:= `Simplify[variance[0, 0, 1, 1, λ01, α01 ]]`

Out[16]= 1

Definition of Jacobian entries

We now define the entries in the (2x2) Jacobian matrix of the mapping g: new mean derived w.r.t. old mean, new mean derived w.r.t. old variance, new variance derived w.r.t.o old mean, new variance derived w.r.t.o old variance.

We derive the mean (first moment) with respect to mu to obtain the entry H11 of the Jacobian:

In[17]:= `Simplify[D[firstMoment[μ, ω, ν, τ, λ, α ], μ]]`

Out[17]=
$$\frac{1}{2}\lambda\,\omega\left(1+\mathrm{Erf}\left[\frac{\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]+e^{\frac{\nu\,\tau}{2}+\mu\,\omega}\,\alpha\,\mathrm{Erfc}\left[\frac{\nu\,\tau+\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]\right)$$

In[18]:= `H11[μ_, ω_, ν_, τ_, λ_, α_ ] :=`
$$\frac{1}{2}\lambda\,\omega\left(1+\mathrm{Erf}\left[\frac{\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]+e^{\frac{\nu\,\tau}{2}+\mu\,\omega}\,\alpha\,\mathrm{Erfc}\left[\frac{\nu\,\tau+\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]\right)$$

We derive the mean (first moment) with respect to nu to obtain the entry H12 of the Jacobian:

In[19]:= `Simplify[D[firstMoment[μ, ω, ν, τ, λ, α ], ν]]`

Out[19]=
$$\frac{1}{4\sqrt{\pi}\,\sqrt{\nu\,\tau}}e^{-\frac{\mu^2\omega^2}{2\,\nu\,\tau}}\lambda\,\tau\left(-\sqrt{2}\,(-1+\alpha)+e^{\frac{(\nu\,\tau+\mu\,\omega)^2}{2\,\nu\,\tau}}\sqrt{\pi}\,\alpha\,\sqrt{\nu\,\tau}\,\mathrm{Erfc}\left[\frac{\nu\,\tau+\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]\right)$$

In[20]:= `H12[μ_, ω_, ν_, τ_, λ_, α_ ] :=`
$$\left(e^{-\frac{\mu^2\omega^2}{2\,\nu\,\tau}}\lambda\,\tau\left(-\sqrt{2}\,(-1+\alpha)+e^{\frac{(\nu\,\tau+\mu\,\omega)^2}{2\,\nu\,\tau}}\sqrt{\pi}\,\alpha\,\sqrt{\nu\,\tau}\,\mathrm{Erfc}\left[\frac{\nu\,\tau+\mu\,\omega}{\sqrt{2}\,\sqrt{\nu\,\tau}}\right]\right)\right)\Big/\left(4\sqrt{\pi}\,\sqrt{\nu\,\tau}\right)$$

We derive the variance with respect to mu to obtain the entry H21 of the Jacobian:

In[21]:= `D[variance[μ, ω, ν, τ, λ, α], μ]`

Out[21]=

$$-\frac{1}{2}\lambda^2\left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}+\mu\omega+\mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-\alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

$$\left(\omega+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\alpha\omega}{\sqrt{\nu\tau}}-\frac{e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\alpha\omega}{\sqrt{\nu\tau}}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\mu\omega^2}{\sqrt{\nu\tau}}-\right.$$

$$\left.\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\mu\sqrt{\nu\tau}\,\omega^2}{\nu\tau}+\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\omega\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)+$$

$$\frac{1}{2}\lambda^2\left(-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\alpha^2\omega}{\sqrt{\nu\tau}}+\frac{2\,e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\alpha^2\omega}{\sqrt{\nu\tau}}-\frac{e^{2\nu\tau+2\mu\omega-\frac{(2\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\alpha^2\omega}{\sqrt{\nu\tau}}+\right.$$

$$e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}\,\omega+2\mu\omega^2-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\mu^2\sqrt{\nu\tau}\,\omega^3}{\nu\tau}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\omega\left(\nu\tau+\mu^2\omega^2\right)}{\sqrt{\nu\tau}}+$$

$$\left.2\mu\omega^2\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-2\,e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\omega\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+2\,e^{2\nu\tau+2\mu\omega}\alpha^2\omega\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

In[22]:= H21[μ_, ω_, ν_, τ_, λ_, α_] :=

$$-\frac{1}{2}\lambda^2\left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}+\mu\omega+\mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-\alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)\left(\omega+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\alpha\omega}{\sqrt{\nu\tau}}-\frac{e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\alpha\omega}{\sqrt{\nu\tau}}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\mu\omega^2}{\sqrt{\nu\tau}}-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\mu\sqrt{\nu\tau}\,\omega^2}{\nu\tau}+\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\omega\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)+$$

$$\frac{1}{2}\lambda^2\left(-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\alpha^2\omega}{\sqrt{\nu\tau}}+\frac{2\,e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\alpha^2\omega}{\sqrt{\nu\tau}}-\frac{e^{2\nu\tau+2\mu\omega-\frac{(2\nu\tau+\mu\omega)^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\alpha^2\omega}{\sqrt{\nu\tau}}+e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}\,\omega+2\mu\omega^2-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\mu^2\sqrt{\nu\tau}\,\omega^3}{\nu\tau}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\,\omega\left(\nu\tau+\mu^2\omega^2\right)}{\sqrt{\nu\tau}}+2\mu\omega^2\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-2\,e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\omega\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+2\,e^{2\nu\tau+2\mu\omega}\alpha^2\omega\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

We derive the variance with respect to nu to obtain the entry H22 of the Jacobian:

In[23]:= `D[variance[μ, ω, ν, τ, λ, α], ν]`

Out[23]=

$$-\frac{1}{2}\lambda^2\left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}+\mu\omega+\mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-\alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

$$\left(\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\tau}{\sqrt{2\pi}\sqrt{\nu\tau}}-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\alpha\mu\tau\omega}{\sqrt{2\pi}(\nu\tau)^{3/2}}-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^2\tau\omega^2}{\sqrt{2\pi}(\nu\tau)^{3/2}}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^2\sqrt{\nu\tau}\,\omega^2}{\sqrt{2\pi}\,\nu^2\tau}-\right.$$

$$\left.\frac{2\,e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha\left(\frac{\tau}{\sqrt{2}\sqrt{\nu\tau}}-\frac{\tau(\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right)}{\sqrt{\pi}}+\frac{1}{2}e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\tau\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)+$$

$$\frac{1}{2}\lambda^2\left(\tau+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\alpha^2\mu\tau\omega}{\sqrt{2\pi}(\nu\tau)^{3/2}}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu\tau\omega}{\sqrt{2\pi}\sqrt{\nu\tau}}+\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^3\sqrt{\nu\tau}\,\omega^3}{\sqrt{2\pi}\,\nu^2\tau}-\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu\tau\omega\left(\nu\tau+\mu^2\omega^2\right)}{\sqrt{2\pi}(\nu\tau)^{3/2}}+\right.$$

$$\frac{4\,e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha^2\left(\frac{\tau}{\sqrt{2}\sqrt{\nu\tau}}-\frac{\tau(\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right)}{\sqrt{\pi}}-\frac{2\,e^{2\nu\tau+2\mu\omega-\frac{(2\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha^2\left(\frac{\sqrt{2}\,\tau}{\sqrt{\nu\tau}}-\frac{\tau(2\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right)}{\sqrt{\pi}}+$$

$$\left.\tau\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]-e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\tau\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]+2\,e^{2\nu\tau+2\mu\omega}\alpha^2\tau\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

In[24]:= H22[μ_, ω_, ν_, τ_, λ_, α_] :=

$$-\frac{1}{2}\lambda^2 \left(e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} + \mu\omega + \mu\omega\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right] - \alpha\,\mathrm{Erfc}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right] + e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)\left(\frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\tau}{\sqrt{2\pi}\sqrt{\nu\tau}} - \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\alpha\mu\tau\omega}{\sqrt{2\pi}(\nu\tau)^{3/2}} - \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^2\tau\omega^2}{\sqrt{2\pi}(\nu\tau)^{3/2}} + \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^2\sqrt{\nu\tau}\,\omega^2}{\sqrt{2\pi}\,\nu^2\tau} - \frac{1}{\sqrt{\pi}}2e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha\left(\frac{\tau}{\sqrt{2}\sqrt{\nu\tau}} - \frac{\tau(\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right) + \frac{1}{2}e^{\frac{\nu\tau}{2}+\mu\omega}\alpha\tau\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right) + \frac{1}{2}\lambda^2\left(\tau + \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\alpha^2\mu\tau\omega}{\sqrt{2\pi}(\nu\tau)^{3/2}} + \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu\tau\omega}{\sqrt{2\pi}\sqrt{\nu\tau}} + \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu^3\sqrt{\nu\tau}\,\omega^3}{\sqrt{2\pi}\,\nu^2\tau} - \frac{e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\mu\tau\omega\left(\nu\tau+\mu^2\omega^2\right)}{\sqrt{2\pi}(\nu\tau)^{3/2}} + \frac{1}{\sqrt{\pi}}4e^{\frac{\nu\tau}{2}+\mu\omega-\frac{(\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha^2\left(\frac{\tau}{\sqrt{2}\sqrt{\nu\tau}} - \frac{\tau(\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right) - \frac{1}{\sqrt{\pi}}2e^{2\nu\tau+2\mu\omega-\frac{(2\nu\tau+\mu\omega)^2}{2\nu\tau}}\alpha^2\left(\frac{\sqrt{2}\,\tau}{\sqrt{\nu\tau}} - \frac{\tau(2\nu\tau+\mu\omega)}{2\sqrt{2}(\nu\tau)^{3/2}}\right) + \tau\,\mathrm{Erf}\left[\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right] - e^{\frac{\nu\tau}{2}+\mu\omega}\alpha^2\tau\,\mathrm{Erfc}\left[\frac{\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right] + 2e^{2\nu\tau+2\mu\omega}\alpha^2\tau\,\mathrm{Erfc}\left[\frac{2\nu\tau+\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right]\right)$$

Jacobian entries for normalized weights

Here are the values of the Jacobian at the fixed point (0,1) for normalized weights.

In[25]:= `N[H11[0, 0, 1, 1, λ01, α01]]`

Out[25]= 0.

In[26]:= `N[H21[0, 0, 1, 1, λ01, α01]]`

Out[26]= 0.

In[27]:= `N[H12[0, 0, 1, 1, λ01, α01]]`

Out[27]= 0.0888348

In[28]:= `N[H22[0, 0, 1, 1, λ01, α01]]`

Out[28]= 0.782648

In[29]:=
```
{{N[H11[0, 0, 1, 1, λ01, α01]], N[H12[0, 0, 1, 1, λ01, α01]]},
  {N[H21[0, 0, 1, 1, λ01, α01]], N[H22[0, 0, 1, 1, λ01, α01]]}} // MatrixForm
```

Out[29]//MatrixForm=

$$\begin{pmatrix} 0. & 0.0888348 \\ 0. & 0.782648 \end{pmatrix}$$

Singular value

We have an explicit formula for the largest singular value of a (2x2) matrix, i.e. the following:

In[30]:=
```
S[μ_, ω_, ν_, τ_, λ_, α_] :=
 0.5 * Sqrt[(H11[μ, ω, ν, τ, λ, α] + H22[μ, ω, ν, τ, λ, α])^2 +
     (H21[μ, ω, ν, τ, λ, α] - H12[μ, ω, ν, τ, λ, α])^2] +
  0.5 * Sqrt[(H11[μ, ω, ν, τ, λ, α] - H22[μ, ω, ν, τ, λ, α])^2 +
     (H21[μ, ω, ν, τ, λ, α] + H12[μ, ω, ν, τ, λ, α])^2]
```

We now check, whether the largest singular value of the Jacobian fixed point (0,1) is smaller then one. If it is smaller than 1, the fixed point (0,1) is a stable fixed point.

In[31]:=
```
N[S[0, 0, 1, 1, λ01, α01]]
```

Out[31]=
```
0.787673
```

Numeric checks for Theorem 1: Singular value < 1; Mapping to domain

For non-normalized weights, the singular value is still below one in certain intervals for mean, variance of activations and mean and variance of weights. We numerically check if the statements of Theorem 1 are true.
First, we check whether the singular value stays below 1.

In[32]:=

In[33]:=
```
Maximize[{S[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

Out[33]=
```
{0.897608, {μ → -0.1, ω → -0.1, ν → 0.8, τ → 1.1}}
```

The maximal value of the singular value is below 1, therefore the numeric check is ok.

```
Maximize[{S[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

We now check whether the mapping stays in the pre-defined intervals.

In[34]:=
```
Minimize[{firstMoment[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

Out[34]=
```
{-0.0310605, {μ → 0.1, ω → -0.1, ν → 0.8, τ → 0.95}}
```

In[35]:=
```
Maximize[{firstMoment[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

Out[35]=
```
{0.0677251, {μ → 0.1, ω → 0.1, ν → 1.5, τ → 1.1}}
```

The new mean is still in the interval [-0.1,0.1], therefore the numeric check is ok.

In[36]:=
```
Minimize[{variance[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

Out[36]=
```
{0.803712, {μ → -0.1, ω → 0.1, ν → 0.8, τ → 0.95}}
```

In[37]:=
```
Maximize[{variance[μ, ω, ν, τ, λ01, α01],
   -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.8 ≤ ν ≤ 1.5 && 0.95 ≤ τ ≤ 1.1}, {μ, ω, ν, τ}]
```

Out[37]=
```
{1.48157, {μ → 0.1, ω → 0.1, ν → 1.5, τ → 1.1}}
```

The new variance is still in the interval [0.8,1.5], therefore the numeric check is ok.

Numeric checks for Theorem 2: new variance smaller than original variance

We can also numerically check theorem 2. We define a new function that represents the difference of the new and the old variance:

```
In[38]:= varianceDifference[μ_, ω_, ν_, τ_, λ_, α_] := variance[μ, ω, ν, τ, λ, α] - ν
```

If this difference is always below zero in the domain with high variance, then the new variance is smaller than the old variance, which is the statement of the theorem:

```
In[39]:= Maximize[{varianceDifference[μ, ω, ν, τ, λ01, α01],
           -1 ≤ μ ≤ 1 && -0.1 ≤ ω ≤ 0.1 && 3 ≤ ν ≤ 16 && 0.8 ≤ τ ≤ 1.25}, {μ, ω, ν, τ}]

Out[39]= {-0.119744, {μ → -1., ω → -0.1, ν → 3., τ → 1.25}}
```

The maximal value of the function is smaller than zero, therefore the numeric check of the statement is ok.

Numeric checks for Theorem 3: new variance larger than original variance

Similarly, we now check the variance difference function in the domains of low variance. If this difference is always above zero in these domains, then the new variance is larger than the old variance, which is the statement of the theorem:

```
In[40]:= Minimize[{varianceDifference[μ, ω, ν, τ, λ01, α01],
           -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.02 ≤ ν ≤ 0.16 && 0.8 ≤ τ ≤ 1.25}, {μ, ω, ν, τ}]

Out[40]= {0.00773618, {μ → 0.0999978, ω → 0.0999978, ν → 0.0200001, τ → 0.800001}}
```

```
In[41]:= Minimize[{varianceDifference[μ, ω, ν, τ, λ01, α01],
           -0.1 ≤ μ ≤ 0.1 && -0.1 ≤ ω ≤ 0.1 && 0.02 ≤ ν ≤ 0.24 && 0.9 ≤ τ ≤ 1.25}, {μ, ω, ν, τ}]

Out[41]= {0.0110293, {μ → -0.1, ω → -0.1, ν → 0.02, τ → 0.9}}
```

The minimal value of these functions is larger than zero, therefore the numeric check of the statement is ok.