Self-Normalizing Neural Networks

Günter Klambauer **Thomas Unterthiner** **Andreas Mayr**

Sepp Hochreiter
LIT AI Lab & Institute of Bioinformatics,
Johannes Kepler University Linz
A-4040 Linz, Austria
{klambauer,unterthiner,mayr,hochreit}@bioinf.jku.at

Abstract

Deep Learning has revolutionized vision via convolutional neural networks (CNNs) and natural language processing via recurrent neural networks (RNNs). However, success stories of Deep Learning with standard feed-forward neural networks (FNNs) are rare. FNNs that perform well are typically shallow and, therefore cannot exploit many levels of abstract representations. We introduce self-normalizing neural networks (SNNs) to enable high-level abstract representations. While batch normalization requires explicit normalization, neuron activations of SNNs automatically converge towards zero mean and unit variance. The activation function of SNNs are "scaled exponential linear units" (SELUs), which induce self-normalizing properties. Using the Banach fixed-point theorem, we prove that activations close to zero mean and unit variance that are propagated through many network layers will converge towards zero mean and unit variance — even under the presence of noise and perturbations. This convergence property of SNNs allows to (1) train deep networks with many layers, (2) employ strong regularization schemes, and (3) to make learning highly robust. Furthermore, for activations not close to unit variance, we prove an upper and lower bound on the variance, thus, vanishing and exploding gradients are impossible. We compared SNNs on (a) 121 tasks from the UCI machine learning repository, on (b) drug discovery benchmarks, and on (c) astronomy tasks with standard FNNs, and other machine learning methods such as random forests and support vector machines. For FNNs we considered (i) ReLU networks without normalization, (ii) batch normalization, (iii) layer normalization, (iv) weight normalization, (v) highway networks, and (vi) residual networks. SNNs significantly outperformed all competing FNN methods at 121 UCI tasks, outperformed all competing methods at the Tox21 dataset, and set a new record at an astronomy data set. The winning SNN architectures are often very deep. Implementations are available at: github.com/bioinf-jku/SNNs

Accepted for publication at NIPS 2017; please cite as:
`Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-Normalizing Neural Networks. In Advances in Neural Information Processing Systems (NIPS).`

Introduction

Deep Learning has set new records at different benchmarks and led to various commercial applications [25, 33]. Recurrent neural networks (RNNs) [18] achieved new levels at speech and natural language

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

processing, for example at the TIMIT benchmark [12] or at language translation [36], and are already employed in mobile devices [31]. RNNs have won handwriting recognition challenges (Chinese and Arabic handwriting) [33, 13, 6] and Kaggle challenges, such as the "Grasp-and Lift EEG" competition. Their counterparts, convolutional neural networks (CNNs) [24] excel at vision and video tasks. CNNs are on par with human dermatologists at the visual detection of skin cancer [9]. The visual processing for self-driving cars is based on CNNs [19], as is the visual input to AlphaGo which has beaten one of the best human GO players [34]. At vision challenges, CNNs are constantly winning, for example at the large ImageNet competition [23, 16], but also almost all Kaggle vision challenges, such as the "Diabetic Retinopathy" and the "Right Whale" challenges [8, 14].

However, looking at Kaggle challenges that are not related to vision or sequential tasks, gradient boosting, random forests, or support vector machines (SVMs) are winning most of the competitions. Deep Learning is notably absent, and for the few cases where FNNs won, they are shallow. For example, the HIGGS challenge, the Merck Molecular Activity challenge, and the Tox21 Data challenge were all won by FNNs with at most four hidden layers. Surprisingly, it is hard to find success stories with FNNs that have many hidden layers, though they would allow for different levels of abstract representations of the input [3].

To robustly train very deep CNNs, batch normalization evolved into a standard to normalize neuron activations to zero mean and unit variance [20]. Layer normalization [2] also ensures zero mean and unit variance, while weight normalization [32] ensures zero mean and unit variance if in the previous layer the activations have zero mean and unit variance. However, training with normalization techniques is perturbed by stochastic gradient descent (SGD), stochastic regularization (like dropout), and the estimation of the normalization parameters. Both RNNs and CNNs can stabilize learning via weight sharing, therefore they are less prone to these perturbations. In contrast, FNNs trained with normalization techniques suffer from these perturbations and have high variance in the training error (see Figure 1). This high variance hinders learning and slows it down. Furthermore, strong regularization, such as dropout, is not possible as it would further increase the variance which in turn would lead to divergence of the learning process. We believe that this sensitivity to perturbations is the reason that FNNs are less successful than RNNs and CNNs.

Self-normalizing neural networks (SNNs) are robust to perturbations and do not have high variance in their training errors (see Figure 1). SNNs push neuron activations to zero mean and unit variance thereby leading to the same effect as batch normalization, which enables to robustly learn many layers. SNNs are based on scaled exponential linear units "SELUs" which induce self-normalizing properties like variance stabilization which in turn avoids exploding and vanishing gradients.

Self-normalizing Neural Networks (SNNs)

Normalization and SNNs. For a neural network with activation function f, we consider two consecutive layers that are connected by a weight matrix $\boldsymbol{W}$. Since the input to a neural network is a random variable, the activations $\boldsymbol{x}$ in the lower layer, the network inputs $\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x}$, and the activations $\boldsymbol{y} = f(\boldsymbol{z})$ in the higher layer are random variables as well. We assume that all activations x_i of the lower layer have mean $\mu := \mathrm{E}(x_i)$ and variance $\nu := \mathrm{Var}(x_i)$. An activation y in the higher layer has mean $\tilde{\mu} := \mathrm{E}(y)$ and variance $\tilde{\nu} := \mathrm{Var}(y)$. Here $\mathrm{E}(.)$ denotes the expectation and $\mathrm{Var}(.)$ the variance of a random variable. A single activation $y = f(z)$ has net input $z = \boldsymbol{w}^T\boldsymbol{x}$. For n units with activation $x_i, 1 \leqslant i \leqslant n$ in the lower layer, we define n times the mean of the weight vector $\boldsymbol{w} \in \mathbb{R}^n$ as $\omega := \sum_{i=1}^n w_i$ and n times the second moment as $\tau := \sum_{i=1}^n w_i^2$.

We consider the mapping g that maps mean and variance of the activations from one layer to mean and variance of the activations in the next layer

$$\begin{pmatrix}\mu\\ \nu\end{pmatrix} \mapsto \begin{pmatrix}\tilde{\mu}\\ \tilde{\nu}\end{pmatrix} : \quad \begin{pmatrix}\tilde{\mu}\\ \tilde{\nu}\end{pmatrix} = g\begin{pmatrix}\mu\\ \nu\end{pmatrix} . \tag{1}$$

Normalization techniques like batch, layer, or weight normalization ensure a mapping g that keeps (μ, ν) and $(\tilde{\mu}, \tilde{\nu})$ close to predefined values, typically $(0, 1)$.

Definition 1 (Self-normalizing neural net). *A neural network is self-normalizing if it possesses a mapping $g : \Omega \mapsto \Omega$ for each activation y that maps mean and variance from one layer to the next*

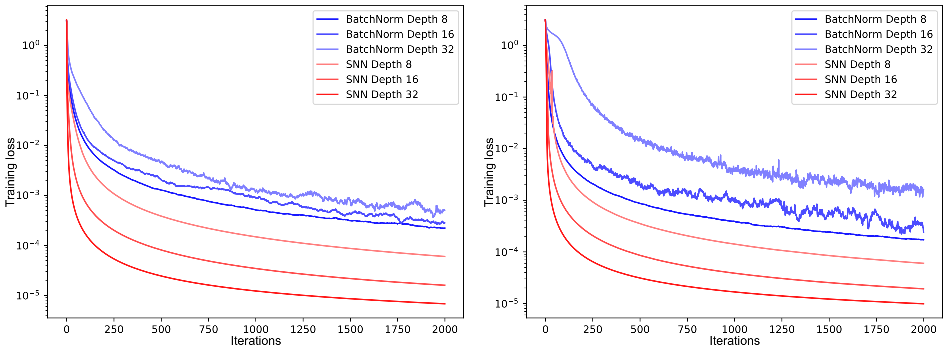

Figure 1: The left panel and the right panel show the training error (y-axis) for feed-forward neural networks (FNNs) with batch normalization (BatchNorm) and self-normalizing networks (SNN) across update steps (x-axis) on the MNIST dataset the CIFAR10 dataset, respectively. We tested networks with 8, 16, and 32 layers and learning rate 1e-5. FNNs with batch normalization exhibit high variance due to perturbations. In contrast, SNNs do not suffer from high variance as they are more robust to perturbations and learn faster.

and has a stable and attracting fixed point depending on (ω, τ) in Ω. Furthermore, the mean and the variance remain in the domain Ω, that is $g(\Omega) \subseteq \Omega$, where $\Omega = \{(\mu, \nu) \mid \mu \in [\mu_{\min}, \mu_{\max}], \nu \in [\nu_{\min}, \nu_{\max}]\}$. When iteratively applying the mapping g, each point within Ω converges to this fixed point.

Therefore, we consider activations of a neural network to be normalized, if both their mean and their variance across samples are within predefined intervals. If mean and variance of $\boldsymbol{x}$ are already within these intervals, then also mean and variance of $\boldsymbol{y}$ remain in these intervals, i.e., the normalization is transitive across layers. Within these intervals, the mean and variance both converge to a fixed point if the mapping g is applied iteratively.

Therefore, SNNs keep normalization of activations when propagating them through layers of the network. The normalization effect is observed across layers of a network: in each layer the activations are getting closer to the fixed point. The normalization effect can also observed be for two fixed layers across learning steps: perturbations of lower layer activations or weights are damped in the higher layer by drawing the activations towards the fixed point. If for all y in the higher layer, ω and τ of the corresponding weight vector are the same, then the fixed points are also the same. In this case we have a unique fixed point for all activations y. Otherwise, in the more general case, ω and τ differ for different y but the mean activations are drawn into $[\mu_{\min}, \mu_{\max}]$ and the variances are drawn into $[\nu_{\min}, \nu_{\max}]$.

Constructing Self-Normalizing Neural Networks. We aim at constructing self-normalizing neural networks by adjusting the properties of the function g. Only two design choices are available for the function g: (1) the activation function and (2) the initialization of the weights.

For the activation function, we propose "scaled exponential linear units" (SELUs) to render a FNN as self-normalizing. The SELU activation function is given by

$$\operatorname{selu}(x) \;=\; \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leqslant 0 \end{cases} . \tag{2}$$

SELUs allow to construct a mapping g with properties that lead to SNNs. SNNs cannot be derived with (scaled) rectified linear units (ReLUs), sigmoid units, $\tanh$ units, and leaky ReLUs. The activation function is required to have (1) negative and positive values for controlling the mean, (2) saturation regions (derivatives approaching zero) to dampen the variance if it is too large in the lower layer, (3) a slope larger than one to increase the variance if it is too small in the lower layer, (4) a continuous curve. The latter ensures a fixed point, where variance damping is equalized by variance increasing. We met these properties of the activation function by multiplying the exponential linear unit (ELU) [7] with $\lambda > 1$ to ensure a slope larger than one for positive net inputs.

For the weight initialization, we propose $\omega = 0$ and $\tau = 1$ for all units in the higher layer. The next paragraphs will show the advantages of this initialization. Of course, during learning these assumptions on the weight vector will be violated. However, we can prove the self-normalizing property even for weight vectors that are not normalized, therefore, the self-normalizing property can be kept during learning and weight changes.

Deriving the Mean and Variance Mapping Function g. We assume that the x_i are independent from each other but share the same mean μ and variance ν. Of course, the independence assumptions is not fulfilled in general. We will elaborate on the independence assumption below. The network input z in the higher layer is $z = \boldsymbol{w}^T \boldsymbol{x}$ for which we can infer the following moments $\mathrm{E}(z) = \sum_{i=1}^{n} w_i \, \mathrm{E}(x_i) = \mu \, \omega$ and $\mathrm{Var}(z) = \mathrm{Var}(\sum_{i=1}^{n} w_i \, x_i) = \nu \, \tau$, where we used the independence of the x_i. The net input z is a weighted sum of independent, but not necessarily identically distributed variables x_i, for which the central limit theorem (CLT) states that z approaches a normal distribution: $z \sim \mathcal{N}(\mu\omega, \sqrt{\nu\tau})$ with density $p_{\mathrm{N}}(z; \mu\omega, \sqrt{\nu\tau})$. According to the CLT, the larger n, the closer is z to a normal distribution. For Deep Learning, broad layers with hundreds of neurons x_i are common. Therefore the assumption that z is normally distributed is met well for most currently used neural networks (see Figure A8). The function g maps the mean and variance of activations in the lower layer to the mean $\tilde{\mu} = \mathrm{E}(y)$ and variance $\tilde{\nu} = \mathrm{Var}(y)$ of the activations y in the next layer:

$$g : \begin{pmatrix} \mu \\ \nu \end{pmatrix} \mapsto \begin{pmatrix} \tilde{\mu} \\ \tilde{\nu} \end{pmatrix} : \quad \begin{aligned} \tilde{\mu}(\mu, \omega, \nu, \tau) \; &= \; \int_{-\infty}^{\infty} \mathrm{selu}(z) \; p_{\mathrm{N}}(z; \mu\omega, \sqrt{\nu\tau}) \, \mathrm{d}z \\ \tilde{\nu}(\mu, \omega, \nu, \tau) \; &= \; \int_{-\infty}^{\infty} \mathrm{selu}(z)^2 \; p_{\mathrm{N}}(z; \mu\omega, \sqrt{\nu\tau}) \, \mathrm{d}z \; - \; (\tilde{\mu})^2 \, . \end{aligned} \tag{3}$$

These integrals can be analytically computed and lead to following mappings of the moments:

$$\begin{aligned} \tilde{\mu} \; = \; & \frac{1}{2}\lambda \left((\mu\omega) \, \mathrm{erf}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) + \right. \\ & \left. \alpha \, e^{\mu\omega + \frac{\nu\tau}{2}} \, \mathrm{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) - \alpha \, \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) + \sqrt{\frac{2}{\pi}} \sqrt{\nu\tau} e^{-\frac{(\mu\omega)^2}{2(\nu\tau)}} + \mu\omega \right) \end{aligned} \tag{4}$$

$$\begin{aligned} \tilde{\nu} \; = \; & \frac{1}{2}\lambda^2 \left(\left((\mu\omega)^2 + \nu\tau \right) \left(2 - \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) \right) + \alpha^2 \left(-2 e^{\mu\omega + \frac{\nu\tau}{2}} \, \mathrm{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) \right. \right. \\ & \left. \left. + e^{2(\mu\omega + \nu\tau)} \, \mathrm{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) \right) + \sqrt{\frac{2}{\pi}} (\mu\omega) \sqrt{\nu\tau} e^{-\frac{(\mu\omega)^2}{2(\nu\tau)}} \right) - (\tilde{\mu})^2 \end{aligned} \tag{5}$$

Stable and Attracting Fixed Point $(0, 1)$ for Normalized Weights. We assume a normalized weight vector $\boldsymbol{w}$ with $\omega = 0$ and $\tau = 1$. Given a fixed point (μ, ν), we can solve equations Eq. (4) and Eq. (5) for α and λ. We chose the fixed point $(\mu, \nu) = (0, 1)$, which is typical for activation normalization. We obtain the fixed point equations $\tilde{\mu} = \mu = 0$ and $\tilde{\nu} = \nu = 1$ that we solve for α and λ and obtain the solutions $\alpha_{01} \approx 1.6733$ and $\lambda_{01} \approx 1.0507$, where the subscript 01 indicates that these are the parameters for fixed point $(0, 1)$. The analytical expressions for α_{01} and λ_{01} are given in Eq. (14). We are interested whether the fixed point $(\mu, \nu) = (0, 1)$ is stable and attracting. If the Jacobian of g has a norm smaller than 1 at the fixed point, then g is a contraction mapping and the fixed point is stable. The (2x2)-Jacobian $\mathcal{J}(\mu, \nu)$ of $g : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\nu})$ evaluated at the fixed point $(0, 1)$ with α_{01} and λ_{01} is

$$\mathcal{J}(\mu, \nu) \; = \; \begin{pmatrix} \partial \frac{\mu^{\mathrm{new}}(\mu,\nu)}{\partial \mu} & \partial \frac{\mu^{\mathrm{new}}(\mu,\nu)}{\partial \nu} \\ \partial \frac{\nu^{\mathrm{new}}(\mu,\nu)}{\partial \mu} & \partial \frac{\nu^{\mathrm{new}}(\mu,\nu)}{\partial \nu} \end{pmatrix} , \qquad \mathcal{J}(0, 1) \; = \; \begin{pmatrix} 0.0 & 0.088834 \\ 0.0 & 0.782648 \end{pmatrix} . \tag{6}$$

The spectral norm of $\mathcal{J}(0, 1)$ (its largest singular value) is $0.7877 < 1$. That means g is a contraction mapping around the fixed point $(0, 1)$ (the mapping is depicted in Figure 2). Therefore, $(0, 1)$ is a stable fixed point of the mapping g.

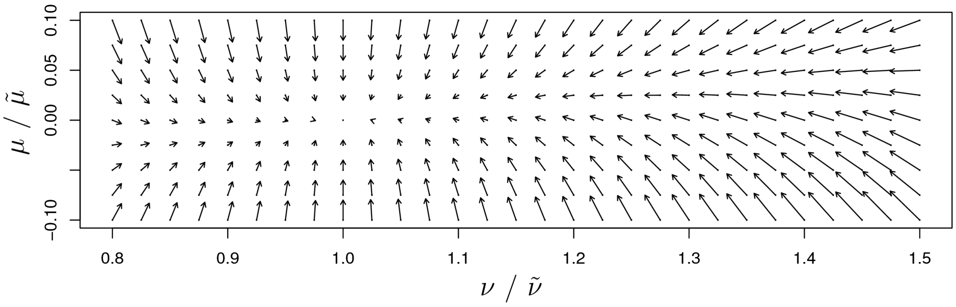

Figure 2: For $\omega = 0$ and $\tau = 1$, the mapping g of mean μ (x-axis) and variance ν (y-axis) to the next layer's mean $\tilde{\mu}$ and variance $\tilde{\nu}$ is depicted. Arrows show in which direction (μ, ν) is mapped by $g : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\nu})$. The fixed point of the mapping g is $(0, 1)$.

Stable and Attracting Fixed Points for Unnormalized Weights. A normalized weight vector $\boldsymbol{w}$ cannot be ensured during learning. For SELU parameters $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$, we show in the next theorem that if (ω, τ) is close to $(0, 1)$, then g still has an attracting and stable fixed point that is close to $(0, 1)$. Thus, in the general case there still exists a stable fixed point which, however, depends on (ω, τ). If we restrict (μ, ν, ω, τ) to certain intervals, then we can show that (μ, ν) is mapped to the respective intervals. Next we present the central theorem of this paper, from which follows that SELU networks are self-normalizing under mild conditions on the weights.

Theorem 1 (Stable and Attracting Fixed Points)**.** *We assume $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. We restrict the range of the variables to the following intervals $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.95, 1.1]$, that define the functions' domain Ω. For $\omega = 0$ and $\tau = 1$, the mapping Eq. (3) has the stable fixed point $(\mu, \nu) = (0, 1)$, whereas for other ω and τ the mapping Eq. (3) has a stable and attracting fixed point depending on (ω, τ) in the (μ, ν)-domain: $\mu \in [-0.03106, 0.06773]$ and $\nu \in [0.80009, 1.48617]$. All points within the (μ, ν)-domain converge when iteratively applying the mapping Eq. (3) to this fixed point.*

Proof. We provide a proof sketch (see detailed proof in Appendix Section A3). With the Banach fixed point theorem we show that there exists a unique attracting and stable fixed point. To this end, we have to prove that a) g is a contraction mapping and b) that the mapping stays in the domain, that is, $g(\Omega) \subseteq \Omega$. The spectral norm of the Jacobian of g can be obtained via an explicit formula for the largest singular value for a 2×2 matrix. g is a contraction mapping if its spectral norm is smaller than 1. We perform a computer-assisted proof to evaluate the largest singular value on a fine grid and ensure the precision of the computer evaluation by an error propagation analysis of the implemented algorithms on the according hardware. Singular values between grid points are upper bounded by the mean value theorem. To this end, we bound the derivatives of the formula for the largest singular value with respect to ω, τ, μ, ν. Then we apply the mean value theorem to pairs of points, where one is on the grid and the other is off the grid. This shows that for all values of ω, τ, μ, ν in the domain Ω, the spectral norm of g is smaller than one. Therefore, g is a contraction mapping on the domain Ω. Finally, we show that the mapping g stays in the domain Ω by deriving bounds on $\tilde{\mu}$ and $\tilde{\nu}$. Hence, the Banach fixed-point theorem holds and there exists a unique fixed point in Ω that is attained. $\square$

Consequently, feed-forward neural networks with many units in each layer and with the SELU activation function are self-normalizing (see definition 1), which readily follows from Theorem 1. To give an intuition, the main property of SELUs is that they damp the variance for negative net inputs and increase the variance for positive net inputs. The variance damping is stronger if net inputs are further away from zero while the variance increase is stronger if net inputs are close to zero. Thus, for large variance of the activations in the lower layer the damping effect is dominant and the variance decreases in the higher layer. Vice versa, for small variance the variance increase is dominant and the variance increases in the higher layer.

However, we cannot guarantee that mean and variance remain in the domain Ω. Therefore, we next treat the case where (μ, ν) are outside Ω. It is especially crucial to consider ν because this variable has much stronger influence than μ. Mapping ν across layers to a high value corresponds to an exploding gradient, since the Jacobian of the activation of high layers with respect to activations in lower layers has large singular values. Analogously, mapping ν across layers to a low value corresponds to an vanishing gradient. Bounding the mapping of ν from above and below would avoid both exploding and vanishing gradients. Theorem 2 states that the variance of neuron activations of SNNs is bounded from above, and therefore ensures that SNNs learn robustly and do not suffer from exploding gradients.

Theorem 2 (Decreasing ν)**.** *For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain Ω^+: $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$, $3 \leqslant \nu \leqslant 16$, and $0.8 \leqslant \tau \leqslant 1.25$, we have for the mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5): $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) < \nu$.*

The proof can be found in the Appendix Section A3. Thus, when mapped across many layers, the variance in the interval $[3, 16]$ is mapped to a value below 3. Consequently, all fixed points (μ, ν) of the mapping g (Eq. (3)) have $\nu < 3$. Analogously, Theorem 3 states that the variance of neuron activations of SNNs is bounded from below, and therefore ensures that SNNs do not suffer from vanishing gradients.

Theorem 3 (Increasing ν)**.** *We consider $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain Ω^-: $-0.1 \leqslant \mu \leqslant 0.1$, and $-0.1 \leqslant \omega \leqslant 0.1$. For the domain $0.02 \leqslant \nu \leqslant 0.16$ and $0.8 \leqslant \tau \leqslant 1.25$ as well as for the domain $0.02 \leqslant \nu \leqslant 0.24$ and $0.9 \leqslant \tau \leqslant 1.25$, the mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5) increases: $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) > \nu$.*

The proof can be found in the Appendix Section A3. All fixed points (μ, ν) of the mapping g (Eq. (3)) ensure for $0.8 \leqslant \tau$ that $\tilde{\nu} > 0.16$ and for $0.9 \leqslant \tau$ that $\tilde{\nu} > 0.24$. Consequently, the variance mapping Eq. (5) ensures a lower bound on the variance ν. Therefore SELU networks control the variance of the activations and push it into an interval, whereafter the mean and variance move toward the fixed point. Thus, SELU networks are steadily normalizing the variance and subsequently normalizing the mean, too. In all experiments, we observed that self-normalizing neural networks push the mean and variance of activations into the domain Ω .

Initialization. Since SNNs have a fixed point at zero mean and unit variance for normalized weights $\omega = \sum_{i=1}^{n} w_i = 0$ and $\tau = \sum_{i=1}^{n} w_i^2 = 1$ (see above), we initialize SNNs such that these constraints are fulfilled in expectation. We draw the weights from a Gaussian distribution with $\mathrm{E}(w_i) = 0$ and variance $\mathrm{Var}(w_i) = 1/n$. Uniform and truncated Gaussian distributions with these moments led to networks with similar behavior. The "MSRA initialization" is similar since it uses zero mean and variance $2/n$ to initialize the weights [17]. The additional factor 2 counters the effect of rectified linear units.

New Dropout Technique. Standard dropout randomly sets an activation x to zero with probability $1 - q$ for $0 < q \leqslant 1$. In order to preserve the mean, the activations are scaled by $1/q$ during training. If x has mean $\mathrm{E}(x) = \mu$ and variance $\mathrm{Var}(x) = \nu$, and the dropout variable d follows a binomial distribution $B(1, q)$, then the mean $\mathrm{E}(1/q d x) = \mu$ is kept. Dropout fits well to rectified linear units, since zero is in the low variance region and corresponds to the default value. For scaled exponential linear units, the default and low variance value is $\lim_{x \to -\infty} \mathrm{selu}(x) = -\lambda\alpha = \alpha'$. Therefore, we propose "alpha dropout", that randomly sets inputs to α'. The new mean and new variance is $\mathrm{E}(xd + \alpha'(1-d)) = q\mu + (1-q)\alpha'$, and $\mathrm{Var}(xd + \alpha'(1-d)) = q((1-q)(\alpha' - \mu)^2 + \nu)$. We aim at keeping mean and variance to their original values after "alpha dropout", in order to ensure the self-normalizing property even for "alpha dropout". The affine transformation $a(xd + \alpha'(1-d)) + b$ allows to determine parameters a and b such that mean and variance are kept to their values: $\mathrm{E}(a(xd + \alpha'(1-d)) + b) = \mu$ and $\mathrm{Var}(a(xd + \alpha'(1-d)) + b) = \nu$. In contrast to dropout, a and b will depend on μ and ν, however our SNNs converge to activations with zero mean and unit variance. With $\mu = 0$ and $\nu = 1$, we obtain $a = \left(q + \alpha'^2 q(1-q)\right)^{-1/2}$ and $b = -\left(q + \alpha'^2 q(1-q)\right)^{-1/2} ((1-q)\alpha')$. The parameters a and b only depend on the dropout rate $1 - q$ and the most negative activation α'. Empirically, we found that dropout rates $1 - q = 0.05$ or 0.10 lead to models with good performance. "Alpha-dropout" fits well to scaled exponential linear units by randomly setting activations to the negative saturation value.

Applicability of the central limit theorem and independence assumption. In the derivative of the mapping (Eq. (3)), we used the central limit theorem (CLT) to approximate the network inputs $z = \sum_{i=1}^{n} w_i x_i$ with a normal distribution. We justified normality because network inputs represent a weighted sum of the inputs x_i, where for Deep Learning n is typically large. The Berry-Esseen theorem states that the convergence rate to normality is $n^{-1/2}$ [22]. In the classical version of the CLT, the random variables have to be independent and identically distributed, which typically does not hold for neural networks. However, the Lyapunov CLT does not require the variable to be identically distributed anymore. Furthermore, even under weak dependence, sums of random variables converge in distribution to a Gaussian distribution [5].

Experiments

We compare SNNs to other deep networks at different benchmarks. Hyperparameters such as number of layers (blocks), neurons per layer, learning rate, and dropout rate, are adjusted by grid-search for each dataset on a separate validation set (see Section A4). We compare the following FNN methods:

- **"MSRAinit":** FNNs without normalization and with ReLU activations and "Microsoft weight initialization" [17].
- **"BatchNorm":** FNNs with batch normalization [20].
- **"LayerNorm":** FNNs with layer normalization [2].
- **"WeightNorm":** FNNs with weight normalization [32].
- **"Highway":** Highway networks [35].
- **"ResNet":** Residual networks [16] adapted to FNNs using residual blocks with 2 or 3 layers with rectangular or diavolo shape.
- **"SNNs":** Self normalizing networks with SELUs with $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ and the proposed dropout technique and initialization strategy.

121 UCI Machine Learning Repository datasets. The benchmark comprises 121 classification datasets from the UCI Machine Learning repository [10] from diverse application areas, such as physics, geology, or biology. The size of the datasets ranges between 10 and $130,000$ data points and the number of features from 4 to 250. In abovementioned work [10], there were methodological mistakes [37] which we avoided here. Each compared FNN method was optimized with respect to its architecture and hyperparameters on a validation set that was then removed from the subsequent analysis. The selected hyperparameters served to evaluate the methods in terms of accuracy on the pre-defined test sets (details on the hyperparameter selection are given in Section A4). The accuracies are reported in the Table A11. We ranked the methods by their accuracy for each prediction task and compared their average ranks. SNNs significantly outperform all competing networks in pairwise comparisons (paired Wilcoxon test across datasets) as reported in Table 1 (left panel).

We further included 17 machine learning methods representing diverse method groups [10] in the comparison and the grouped the data sets into "small" and "large" data sets (for details see Section A4). On 75 small datasets with less than 1000 data points, random forests and SVMs outperform SNNs and other FNNs. On 46 larger datasets with at least 1000 data points, SNNs show the highest performance followed by SVMs and random forests (see right panel of Table 1, for complete results see Tables A12 and A12). Overall, SNNs have outperformed state of the art machine learning methods on UCI datasets with more than 1,000 data points.

Typically, hyperparameter selection chose SNN architectures that were much deeper than the selected architectures of other FNNs, with an average depth of 10.8 layers, compared to average depths of 6.0 for BatchNorm, 3.8 WeightNorm, 7.0 LayerNorm, 5.9 Highway, and 7.1 for MSRAinit networks. For ResNet, the average number of blocks was 6.35. SNNs with many more than 4 layers often provide the best predictive accuracies across all neural networks.

Drug discovery: The Tox21 challenge dataset. The Tox21 challenge dataset comprises about 12,000 chemical compounds whose twelve toxic effects have to be predicted based on their chemical

Table 1: **Left:** Comparison of seven FNNs on 121 UCI tasks. We consider the average rank difference to rank 4, which is the average rank of seven methods with random predictions. The first column gives the method, the second the average rank difference, and the last the p-value of a paired Wilcoxon test whether the difference to the best performing method is significant. SNNs significantly outperform all other methods. **Right:** Comparison of 24 machine learning methods (ML) on the UCI datasets with more than 1000 data points. The first column gives the method, the second the average rank difference to rank 12.5, and the last the p-value of a paired Wilcoxon test whether the difference to the best performing method is significant. Methods that were significantly worse than the best method are marked with "*". The full tables can be found in Table A11, Table A12 and Table A13. SNNs outperform all competing methods.

FNN method comparison			ML method comparison		
Method	avg. rank diff.	p-value	Method	avg. rank diff.	p-value
SNN	-0.756		SNN	-6.7	
MSRAinit	-0.240*	2.7e-02	SVM	-6.4	5.8e-01
LayerNorm	-0.198*	1.5e-02	RandomForest	-5.9	2.1e-01
Highway	0.021*	1.9e-03	MSRAinit	-5.4*	4.5e-03
ResNet	0.273*	5.4e-04	LayerNorm	-5.3	7.1e-02
WeightNorm	0.397*	7.8e-07	Highway	-4.6*	1.7e-03
BatchNorm	0.504*	3.5e-06	...	...	...

structure. We used the validation sets of the challenge winners for hyperparameter selection (see Section A4) and the challenge test set for performance comparison. We repeated the whole evaluation procedure 5 times to obtain error bars. The results in terms of average AUC are given in Table 2. In 2015, the challenge organized by the US NIH was won by an ensemble of shallow ReLU FNNs which achieved an AUC of 0.846 [28]. Besides FNNs, this ensemble also contained random forests and SVMs. Single SNNs came close with an AUC of 0.845$\pm$0.003. The best performing SNNs have 8 layers, compared to the runner-ups ReLU networks with layer normalization with 2 and 3 layers. Also batchnorm and weightnorm networks, typically perform best with shallow networks of 2 to 4 layers (Table 2). The deeper the networks, the larger the difference in performance between SNNs and other methods (see columns 5–8 of Table 2). The best performing method is an SNN with 8 layers.

Table 2: Comparison of FNNs at the Tox21 challenge dataset in terms of AUC. The rows represent different methods and the columns different network depth and for ResNets the number of residual blocks ("na": 32 blocks were omitted due to computational constraints). The deeper the networks, the more prominent is the advantage of SNNs. The best networks are SNNs with 8 layers.

	#layers / #blocks						
method	2	3	4	6	8	16	32
SNN	*83.7* ± 0.3	**84.4** ± 0.5	**84.2** ± 0.4	**83.9** ± 0.5	**84.5** ± 0.2	**83.5** ± 0.5	**82.5** ± 0.7
Batchnorm	80.0 ± 0.5	79.8 ± 1.6	77.2 ± 1.1	77.0 ± 1.7	75.0 ± 0.9	73.7 ± 2.0	76.0 ± 1.1
WeightNorm	*83.7* ± 0.8	82.9 ± 0.8	82.2 ± 0.9	*82.5* ± 0.6	*81.9* ± 1.2	78.1 ± 1.3	56.6 ± 2.6
LayerNorm	**84.3** ± 0.3	*84.3* ± 0.5	*84.0* ± 0.2	*82.5* ± 0.8	80.9 ± 1.8	78.7 ± 2.3	78.8 ± 0.8
Highway	83.3 ± 0.9	83.0 ± 0.5	82.6 ± 0.9	82.4 ± 0.8	80.3 ± 1.4	80.3 ± 2.4	79.6 ± 0.8
MSRAinit	82.7 ± 0.4	81.6 ± 0.9	81.1 ± 1.7	80.6 ± 0.6	80.9 ± 1.1	80.2 ± 1.1	80.4 ± 1.9
ResNet	82.2 ± 1.1	80.0 ± 2.0	80.5 ± 1.2	81.2 ± 0.7	81.8 ± 0.6	*81.2* ± 0.6	na

Astronomy: Prediction of pulsars in the HTRU2 dataset. Since a decade, machine learning methods have been used to identify pulsars in radio wave signals [27]. Recently, the High Time Resolution Universe Survey (HTRU2) dataset has been released with 1,639 real pulsars and 16,259 spurious signals. Currently, the highest AUC value of a 10-fold cross-validation is 0.976 which has been achieved by Naive Bayes classifiers followed by decision tree C4.5 with 0.949 and SVMs with 0.929. We used eight features constructed by the PulsarFeatureLab as used previously [27]. We assessed the performance of FNNs using 10-fold nested cross-validation, where the hyperparameters were selected in the inner loop on a validation set (for details on the hyperparameter selection see

Section A4). Table 3 reports the results in terms of AUC. SNNs outperform all other methods and have pushed the state-of-the-art to an AUC of 0.98.

Table 3: Comparison of FNNs and reference methods at HTRU2 in terms of AUC. The first, fourth and seventh column give the method, the second, fifth and eight column the AUC averaged over 10 cross-validation folds, and the third and sixth column the p-value of a paired Wilcoxon test of the AUCs against the best performing method across the 10 folds. FNNs achieve better results than Naive Bayes (NB), C4.5, and SVM. SNNs exhibit the best performance and set a new record.

	FNN methods			FNN methods		ref. methods	
method	AUC	p-value	method	AUC	p-value	method	AUC
SNN	0.9803 $\pm$ 0.010						
MSRAinit	0.9791 $\pm$ 0.010	3.5e-01	LayerNorm	0.9762* $\pm$ 0.011	1.4e-02	NB	0.976
WeightNorm	0.9786* $\pm$ 0.010	2.4e-02	BatchNorm	0.9760 $\pm$ 0.013	6.5e-02	C4.5	0.946
Highway	0.9766* $\pm$ 0.009	9.8e-03	ResNet	0.9753* $\pm$ 0.010	6.8e-03	SVM	0.929

Conclusion

We have introduced self-normalizing neural networks for which we have proved that neuron activations are pushed towards zero mean and unit variance when propagated through the network. Additionally, for activations not close to unit variance, we have proved an upper and lower bound on the variance mapping. Consequently, SNNs do not face vanishing and exploding gradient problems. Therefore, SNNs work well for architectures with many layers, allowed us to introduce a novel regularization scheme, and learn very robustly. On 121 UCI benchmark datasets, SNNs have outperformed other FNNs with and without normalization techniques, such as batch, layer, and weight normalization, or specialized architectures, such as Highway or Residual networks. SNNs also yielded the best results on drug discovery and astronomy tasks. The best performing SNN architectures are typically very deep in contrast to other FNNs.

Acknowledgments

This work was supported by IWT research grant IWT150865 (Exaptation), H2020 project grant 671555 (ExCAPE), grant IWT135122 (ChemBioBridge), Zalando SE with Research Agreement 01/2016, Audi.JKU Deep Learning Center, Audi Electronic Venture GmbH, and the NVIDIA Corporation.

References

The references are provided in Section A7.

Appendix

Contents

This appendix is organized as follows: the first section sets the background, definitions, and formulations. The main theorems are presented in the next section. The following section is devoted to the proofs of these theorems. The next section reports additional results and details on the performed computational experiments, such as hyperparameter selection. The last section shows that our theoretical bounds can be confirmed by numerical methods as a sanity check.

The proof of theorem 1 is based on the Banach's fixed point theorem for which we require (1) a contraction mapping, which is proved in Subsection A3.4.1 and (2) that the mapping stays within its domain, which is proved in Subsection A3.4.2 For part (1), the proof relies on the main Lemma 12, which is a computer-assisted proof, and can be found in Subsection A3.4.1. The validity of the computer-assisted proof is shown in Subsection A3.4.5 by error analysis and the precision of the functions' implementation. The last Subsection A3.4.6 compiles various lemmata with intermediate results that support the proofs of the main lemmata and theorems.

A1 Background

We consider a neural network with **activation function** f and two consecutive layers that are connected by **weight matrix** $\boldsymbol{W}$. Since samples that serve as input to the neural network are chosen according to a distribution, the **activations** $\boldsymbol{x}$ **in the lower layer**, the **network inputs** $\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x}$, and **activations** $\boldsymbol{y} = f(\boldsymbol{z})$ **in the higher layer** are all random variables. We assume that all units x_i in the lower layer have **mean activation** $\mu := \mathrm{E}(x_i)$ and **variance of the activation** $\nu := \mathrm{Var}(x_i)$ and a unit y in the higher layer has mean activation $\tilde{\mu} := \mathrm{E}(y)$ and variance $\tilde{\nu} := \mathrm{Var}(y)$. Here $\mathrm{E}(.)$ denotes the expectation and $\mathrm{Var}(.)$ the variance of a random variable. For activation of unit y, we have net input $z = \boldsymbol{w}^T\boldsymbol{x}$ and the **scaled exponential linear unit (SELU)** activation $y = \mathrm{selu}(z)$, with

$$\mathrm{selu}(x) \ = \ \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leqslant 0 \end{cases} . \tag{7}$$

For n units $x_i, 1 \leqslant i \leqslant n$ in the lower layer and the **weight vector** $\boldsymbol{w} \in \mathbb{R}^n$, we define n **times the mean** by $\omega := \sum_{i=1}^n w_i$ and n **times the second moment** by $\tau := \sum_{i=1}^n w_i^2$.

We define a **mapping** g from mean μ and variance ν of one layer to the mean $\tilde{\mu}$ and variance $\tilde{\nu}$ in the next layer:

$$g : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\nu}) . \tag{8}$$

For neural networks with scaled exponential linear units, the mean is of the activations in the next layer computed according to

$$\tilde{\mu} \ = \ \int_{-\infty}^{0} \lambda\alpha(\exp(z) - 1)p_{\mathrm{Gauss}}(z; \mu\omega, \sqrt{\nu\tau})dz \ + \ \int_{0}^{\infty} \lambda z p_{\mathrm{Gauss}}(z; \mu\omega, \sqrt{\nu\tau})dz \ , \tag{9}$$

and the second moment of the activations in the next layer is computed according to

$$\tilde{\xi} \ = \ \int_{-\infty}^{0} \lambda^2\alpha^2(\exp(z) - 1)^2 p_{\mathrm{Gauss}}(z; \mu\omega, \sqrt{\nu\tau})dz \ + \ \int_{0}^{\infty} \lambda^2 z^2 p_{\mathrm{Gauss}}(z; \mu\omega, \sqrt{\nu\tau})dz \ . \tag{10}$$

Therefore, the expressions $\tilde{\mu}$ and $\tilde{\nu}$ have the following form:

$$\begin{aligned}
\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha) \ &= \frac{1}{2}\lambda \left(-(\alpha + \mu\omega)\,\mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + \right. & (11)\\
&\left. \alpha e^{\mu\omega + \frac{\nu\tau}{2}}\,\mathrm{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}e^{-\frac{\mu^2\omega^2}{2\nu\tau}} + 2\mu\omega \right) \\
\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha) \ &= \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha) - (\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha))^2 & (12)\\
\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha) \ &= \ \frac{1}{2}\lambda^2 \left(\left((\mu\omega)^2 + \nu\tau\right)\left(\mathrm{erf}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 1\right) + \right. & (13)\\
&\alpha^2 \left(-2e^{\mu\omega + \frac{\nu\tau}{2}}\,\mathrm{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + e^{2(\mu\omega + \nu\tau)}\,\mathrm{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \right. \\
&\left. \left. \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \sqrt{\frac{2}{\pi}}(\mu\omega)\sqrt{\nu\tau}e^{-\frac{(\mu\omega)^2}{2(\nu\tau)}} \right)
\end{aligned}$$

We solve equations Eq. 4 and Eq. 5 for fixed points $\tilde{\mu} = \mu$ and $\tilde{\nu} = \nu$. For a normalized weight vector with $\omega = 0$ and $\tau = 1$ and the **fixed point** $(\mu, \nu) = (0, 1)$, we can solve equations Eq. 4 and Eq. 5 for α and λ. We denote the solutions to fixed point $(\mu, \nu) = (0, 1)$ by α_{01} and λ_{01}.

$$\begin{aligned}
\alpha_{01} &= -\frac{\sqrt{\frac{2}{\pi}}}{\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\right)\exp\left(\frac{1}{2}\right) - 1} \approx 1.67326 & (14)\\
\lambda_{01} &= \left(1 - \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\right)\sqrt{e}\right)\sqrt{2\pi}
\end{aligned}$$

$$\left(2\operatorname{erfc}\left(\sqrt{2}\right)e^2+\pi\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\right)^2 e-2(2+\pi)\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\right)\sqrt{e}+\pi+2\right)^{-1/2}$$

$$\lambda_{01} \approx 1.0507 \ .$$

The parameters α_{01} and λ_{01} ensure

$$\tilde{\mu}(0,0,1,1,\lambda_{01},\alpha_{01}) = 0$$
$$\tilde{\nu}(0,0,1,1,\lambda_{01},\alpha_{01}) = 1$$

Since we focus on the fixed point $(\mu, \nu) = (0, 1)$, we assume throughout the analysis that $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. We consider the functions $\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})$, $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})$, and $\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})$ on the **domain** $\Omega = \{(\mu, \omega, \nu, \tau) \mid \mu \in [\mu_{\min}, \mu_{\max}] = [-0.1, 0.1], \omega \in [\omega_{\min}, \omega_{\max}] = [-0.1, 0.1], \nu \in [\nu_{\min}, \nu_{\max}] = [0.8, 1.5], \tau \in [\tau_{\min}, \tau_{\max}] = [0.95, 1.1]\}$.

Figure 2 visualizes the mapping g for $\omega = 0$ and $\tau = 1$ and α_{01} and λ_{01} at few pre-selected points. It can be seen that $(0, 1)$ is an attracting fixed point of the mapping g.

A2 Theorems

A2.1 Theorem 1: Stable and Attracting Fixed Points Close to (0,1)

Theorem 1 shows that the mapping g defined by Eq. (4) and Eq. (5) exhibits a stable and attracting fixed point close to zero mean and unit variance. Theorem 1 establishes the self-normalizing property of self-normalizing neural networks (SNNs). The stable and attracting fixed point leads to robust learning through many layers.

Theorem 1 (Stable and Attracting Fixed Points)**.** *We assume $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. We restrict the range of the variables to the domain $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.95, 1.1]$. For $\omega = 0$ and $\tau = 1$, the mapping Eq. (4) and Eq. (5) has the stable fixed point $(\mu, \nu) = (0, 1)$. For other ω and τ the mapping Eq. (4) and Eq. (5) has a stable and attracting fixed point depending on (ω, τ) in the (μ, ν)-domain: $\mu \in [-0.03106, 0.06773]$ and $\nu \in [0.80009, 1.48617]$. All points within the (μ, ν)-domain converge when iteratively applying the mapping Eq. (4) and Eq. (5) to this fixed point.*

A2.2 Theorem 2: Decreasing Variance from Above

The next Theorem 2 states that the variance of unit activations does not explode through consecutive layers of self-normalizing networks. Even more, a large variance of unit activations decreases when propagated through the network. In particular this ensures that exploding gradients will never be observed. In contrast to the domain in previous subsection, in which $\nu \in [0.8, 1.5]$, we now consider a domain in which the variance of the inputs is higher $\nu \in [3, 16]$ and even the range of the mean is increased $\mu \in [-1, 1]$. We denote this new domain with the symbol Ω^{++} to indicate that the variance lies above the variance of the original domain Ω. In Ω^{++}, we can show that the variance $\tilde{\nu}$ in the next layer is always smaller then the original variance ν. Concretely, this theorem states that:

Theorem 2 (Decreasing ν)**.** *For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain Ω^{++}: $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$, $3 \leqslant \nu \leqslant 16$, and $0.8 \leqslant \tau \leqslant 1.25$ we have for the mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5)*

$$\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \ < \ \nu \ . \tag{15}$$

The variance decreases in $[3, 16]$ and all fixed points (μ, ν) of mapping Eq. (5) and Eq. (4) have $\nu < 3$.

A2.3 Theorem 3: Increasing Variance from Below

The next Theorem 3 states that the variance of unit activations does not vanish through consecutive layers of self-normalizing networks. Even more, a small variance of unit activations increases when propagated through the network. In particular this ensures that vanishing gradients will never be observed. In contrast to the first domain, in which $\nu \in [0.8, 1.5]$, we now consider two domains Ω_1^- and Ω_2^- in which the variance of the inputs is lower $0.05 \leqslant \nu \leqslant 0.16$ and $0.05 \leqslant \nu \leqslant 0.24$, and even the parameter τ is different $0.9 \leqslant \tau \leqslant 1.25$ to the original Ω. We denote this new domain with the symbol Ω_i^- to indicate that the variance lies below the variance of the original domain Ω. In Ω_1^- and Ω_2^-, we can show that the variance $\tilde{\nu}$ in the next layer is always larger then the original variance ν, which means that the variance does not vanish through consecutive layers of self-normalizing networks. Concretely, this theorem states that:

Theorem 3 (Increasing ν)**.** *We consider $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the two domains $\Omega_1^- = \{(\mu, \omega, \nu, \tau) \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.16, 0.8 \leqslant \tau \leqslant 1.25\}$ and $\Omega_2^- = \{(\mu, \omega, \nu, \tau) \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.24, 0.9 \leqslant \tau \leqslant 1.25\}$.*

The mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5) increases

$$\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \; > \; \nu \tag{16}$$

in both Ω_1^- and Ω_2^-. All fixed points (μ, ν) of mapping Eq. (5) and Eq. (4) ensure for $0.8 \leqslant \tau$ that $\tilde{\nu} > 0.16$ and for $0.9 \leqslant \tau$ that $\tilde{\nu} > 0.24$. Consequently, the variance mapping Eq. (5) and Eq. (4) ensures a lower bound on the variance ν.

A3 Proofs of the Theorems

A3.1 Proof of Theorem 1

We have to show that the mapping g defined by Eq. (4) and Eq. (5) has a stable and attracting fixed point close to $(0, 1)$. To proof this statement and Theorem 1, we apply the Banach fixed point theorem which requires (1) that g is a contraction mapping and (2) that g does not map outside the function's domain, concretely:

Theorem 4 (Banach Fixed Point Theorem)**.** *Let (X, d) be a non-empty complete metric space with a contraction mapping $f : X \to X$. Then f has a unique fixed-point $x_f \in X$ with $f(x_f) = x_f$. Every sequence $x_n = f(x_{n-1})$ with starting element $x_0 \in X$ converges to the fixed point: $x_n \xrightarrow[n \to \infty]{} x_f$.*

Contraction mappings are functions that map two points such that their distance is decreasing:

Definition 2 (Contraction mapping)**.** *A function $f : X \to X$ on a metric space X with distance d is a contraction mapping, if there is a $0 \leqslant \delta < 1$, such that for all points $\boldsymbol{u}$ and $\boldsymbol{v}$ in X: $d(f(\boldsymbol{u}), f(\boldsymbol{v})) \leqslant \delta d(\boldsymbol{u}, \boldsymbol{v})$.*

To show that g is a contraction mapping in Ω with distance $\|.\|_2$, we use the Mean Value Theorem for $u, v \in \Omega$

$$\|g(\boldsymbol{u}) - g(\boldsymbol{v})\|_2 \leqslant M \; \|\boldsymbol{u} - \boldsymbol{v}\|_2, \tag{17}$$

in which M is an upper bound on the spectral norm the Jacobian $\mathcal{H}$ of g. The spectral norm is given by the largest singular value of the Jacobian of g. If the largest singular value of the Jacobian is smaller than 1, the mapping g of the mean and variance to the mean and variance in the next layer is contracting. We show that the largest singular value is smaller than 1 by evaluating the function for the singular value $S(\mu, \omega, \nu, \tau, \lambda, \alpha)$ on a grid. Then we use the Mean Value Theorem to bound the deviation of the function S between grid points. To this end, we have to bound the gradient of S with respect to (μ, ω, ν, τ). If all function values plus gradient times the deltas (differences between grid points and evaluated points) is still smaller than 1, then we have proofed that the function is below 1 (Lemma 12). To show that the mapping does not map outside the function's domain, we derive bounds on the expressions for the mean and the variance (Lemma 13). Section A3.4.1 and Section A3.4.2 are concerned with the contraction mapping and the image of the function domain of g, respectively.

With the results that the largest singular value of the Jacobian is smaller than one (Lemma 12) and that the mapping stays in the domain Ω (Lemma 13), we can prove Theorem 1. We first recall Theorem 1:

Theorem (Stable and Attracting Fixed Points)**.** *We assume $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. We restrict the range of the variables to the domain $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.95, 1.1]$. For $\omega = 0$ and $\tau = 1$, the mapping Eq. (4) and Eq. (5) has the stable fixed point $(\mu, \nu) = (0, 1)$. For other ω and τ the mapping Eq. (4) and Eq. (5) has a stable and attracting fixed point depending on (ω, τ) in the (μ, ν)-domain: $\mu \in [-0.03106, 0.06773]$ and $\nu \in [0.80009, 1.48617]$. All points within the (μ, ν)-domain converge when iteratively applying the mapping Eq. (4) and Eq. (5) to this fixed point.*

Proof. According to Lemma 12 the mapping g (Eq. (4) and Eq. (5)) is a contraction mapping in the given domain, that is, it has a Lipschitz constant smaller than one. We showed that $(\mu, \nu) = (0, 1)$ is a fixed point of the mapping for $(\omega, \tau) = (0, 1)$.

The domain is compact (bounded and closed), therefore it is a complete metric space. We further have to make sure the mapping g does not map outside its domain Ω. According to Lemma 13, the mapping maps into the domain $\mu \in [-0.03106, 0.06773]$ and $\nu \in [0.80009, 1.48617]$.

Now we can apply the Banach fixed point theorem given in Theorem 4 from which the statement of the theorem follows. $\square$

A3.2 Proof of Theorem 2

First we recall Theorem 2:

Theorem (Decreasing ν)**.** *For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain Ω^{++}: $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$, $3 \leqslant \nu \leqslant 16$, and $0.8 \leqslant \tau \leqslant 1.25$ we have for the mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5)*

$$\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \ < \ \nu \ . \tag{18}$$

The variance decreases in $[3, 16]$ and all fixed points (μ, ν) of mapping Eq. (5) and Eq. (4) have $\nu < 3$.

Proof. We start to consider an even larger domain $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$, $1.5 \leqslant \nu \leqslant 16$, and $0.8 \leqslant \tau \leqslant 1.25$. We prove facts for this domain and later restrict to $3 \leqslant \nu \leqslant 16$, i.e. Ω^{++}. We consider the function g of the difference between the second moment $\tilde{\xi}$ in the next layer and the variance ν in the lower layer:

$$g(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \ = \ \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \ - \ \nu \ . \tag{19}$$

If we can show that $g(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) < 0$ for all $(\mu, \omega, \nu, \tau) \in \Omega^{++}$, then we would obtain our desired result $\tilde{\nu} \leqslant \tilde{\xi} < \nu$. The derivative with respect to ν is according to Theorem 16

$$\frac{\partial}{\partial \nu} g(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \ = \ \frac{\partial}{\partial \nu} \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \ - \ 1 \ < \ 0 \ . \tag{20}$$

Therefore g is strictly monotonically decreasing in ν. Since $\tilde{\xi}$ is a function in $\nu\tau$ (these variables only appear as this product), we have for $x = \nu\tau$

$$\frac{\partial}{\partial \nu} \tilde{\xi} \ = \ \frac{\partial}{\partial x} \tilde{\xi} \, \frac{\partial x}{\partial \nu} \ = \ \frac{\partial}{\partial x} \tilde{\xi} \, \tau \tag{21}$$

and

$$\frac{\partial}{\partial \tau} \tilde{\xi} \ = \ \frac{\partial}{\partial x} \tilde{\xi} \, \frac{\partial x}{\partial \tau} \ = \ \frac{\partial}{\partial x} \tilde{\xi} \, \nu \ . \tag{22}$$

Therefore we have according to Theorem 16:

$$\frac{\partial}{\partial \tau} \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \ = \ \frac{\nu}{\tau} \, \frac{\partial}{\partial \nu} \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \ > \ 0 \ . \tag{23}$$

Therefore

$$\frac{\partial}{\partial \tau} g(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \ = \ \frac{\partial}{\partial \tau} \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \ > \ 0 \ . \tag{24}$$

Consequently, g is strictly monotonically increasing in τ. Now we consider the derivative with respect to μ and ω. We start with $\frac{\partial}{\partial \mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, which is

$$\begin{aligned}
&\frac{\partial}{\partial \mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha) \; = \qquad (25)\\
&\lambda^2\omega\left(\alpha^2\left(-e^{\mu\omega+\frac{\nu\tau}{2}}\right)\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)+\right.\\
&\left.\alpha^2 e^{2\mu\omega+2\nu\tau}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)+\mu\omega\left(2-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)+\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right) .
\end{aligned}$$

We consider the sub-function

$$\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}-\alpha^2\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) . \qquad (26)$$

We set $x=\nu\tau$ and $y=\mu\omega$ and obtain

$$\sqrt{\frac{2}{\pi}}\sqrt{x}-\alpha^2\left(e^{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-e^{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right) . \qquad (27)$$

The derivative to this sub-function with respect to y is

$$\begin{aligned}
&\frac{\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}(2x+y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)-e^{\frac{(x+y)^2}{2x}}(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right)}{x} = \qquad (28)\\
&\frac{\sqrt{2}\alpha^2\sqrt{x}\left(\frac{e^{\frac{(2x+y)^2}{2x}}(2x+y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}-\frac{e^{\frac{(x+y)^2}{2x}}(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)}{x} \; > \; 0 .
\end{aligned}$$

The inequality follows from Lemma 24, which states that $ze^{z^2}\operatorname{erfc}(z)$ is monotonically increasing in z. Therefore the sub-function is increasing in y. The derivative to this sub-function with respect to x is

$$\begin{aligned}
&\frac{1}{2\sqrt{\pi}x^2}\sqrt{\pi}\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right. \qquad (29)\\
&\left.-e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right)-\sqrt{2}\left(\alpha^2-1\right)x^{3/2}.
\end{aligned}$$

The sub-function is increasing in x, since the derivative is larger than zero:

$$\begin{aligned}
&\frac{\sqrt{\pi}\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)-e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right)-\sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} \geqslant \qquad (30)\\
&\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2+2}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2+\frac{4}{\pi}}\right)}\right)-\sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} =\\
&\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2\left(\sqrt{2}\sqrt{x}\right)}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+4x}\right)}-\frac{(x-y)(x+y)2\left(\sqrt{2}\sqrt{x}\right)}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+\frac{8x}{\pi}}\right)}\right)-\sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} =\\
&\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+4x}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+\frac{8x}{\pi}}\right)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} >
\end{aligned}$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+2(2x+y)+1}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+0.878\cdot 2(x+y)+0.878^2}\right)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y+1)^2}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y+0.878)^2}\right)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}(2(2x+y)+1)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}(2(x+y)+0.878)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2(x+y)+0.878)(2x-y)(2x+y)2}{\sqrt{\pi}}-\frac{(x-y)(x+y)(2(2x+y)+1)2}{\sqrt{\pi}}\right)}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}x^{3/2}} +
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(-x\left(\alpha^2-1\right)(2(2x+y)+1)(2(x+y)+0.878)\right)}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{8x^3+12x^2y+4.14569x^2+4xy^2-6.76009xy-1.58023x+0.683154y^2}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}x^{3/2}} >
$$

$$
\frac{8x^3-0.1\cdot 12x^2+4.14569x^2+4\cdot(0.0)^2x-6.76009\cdot 0.1x-1.58023x+0.683154\cdot(0.0)^2}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{8x^2+2.94569x-2.25624}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}\sqrt{x}} =
$$

$$
\frac{8(x-0.377966)(x+0.746178)}{(2(2x+y)+1)(2(x+y)+0.878)\sqrt{2}\sqrt{\pi}\sqrt{x}} > 0 .
$$

We explain this chain of inequalities:

- First inequality: We applied Lemma 22 two times.
- Equalities factor out $\sqrt{2}\sqrt{x}$ and reformulate.
- Second inequality part 1: we applied

$$
0 < 2y \Longrightarrow (2x+y)^2+4x+1 < (2x+y)^2+2(2x+y)+1 = (2x+y+1)^2 . \tag{31}
$$

- Second inequality part 2: we show that for $a=\frac{1}{10}\left(\sqrt{\frac{960+169\pi}{\pi}}-13\right)$ following holds: $\frac{8x}{\pi}-\left(a^2+2a(x+y)\right)\geqslant 0$. We have $\frac{\partial}{\partial x}\frac{8x}{\pi}-\left(a^2+2a(x+y)\right)=\frac{8}{\pi}-2a>0$ and $\frac{\partial}{\partial y}\frac{8x}{\pi}-\left(a^2+2a(x+y)\right)=-2a<0$. Therefore the minimum is at border for minimal x and maximal y:

$$
\frac{8\cdot 1.2}{\pi}-\left(\frac{2}{10}\left(\sqrt{\frac{960+169\pi}{\pi}}-13\right)(1.2+0.1)+\left(\frac{1}{10}\left(\sqrt{\frac{960+169\pi}{\pi}}-13\right)\right)^2\right) = 0 . \tag{32}
$$

Thus

$$
\frac{8x}{\pi} \geqslant a^2+2a(x+y) . \tag{33}
$$

for $a=\frac{1}{10}\left(\sqrt{\frac{960+169\pi}{\pi}}-13\right)>0.878$.

- Equalities only solve square root and factor out the resulting terms $(2(2x+y)+1)$ and $(2(x+y)+0.878)$.
- We set $\alpha=\alpha_{01}$ and multiplied out. Thereafter we also factored out x in the numerator. Finally a quadratic equations was solved.

The sub-function has its minimal value for minimal $x = \nu\tau = 1.5 \cdot 0.8 = 1.2$ and minimal $y = \mu\omega = -1 \cdot 0.1 = -0.1$. We further minimize the function

$$\mu\omega e^{\frac{\mu^2\omega^2}{2\nu\tau}} \left(2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \; > \; -0.1 e^{\frac{0.1^2}{2 \cdot 1.2}} \left(2 - \operatorname{erfc}\left(\frac{0.1}{\sqrt{2}\sqrt{1.2}}\right)\right) . \tag{34}$$

We compute the minimum of the term in brackets of $\frac{\partial}{\partial \mu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ in Eq. (25):

$$\begin{aligned}
&\mu\omega e^{\frac{\mu^2\omega^2}{2\nu\tau}} \left(2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \qquad (35)\\
&\alpha_{01}^2 \left(-\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} \; > \\
&\alpha_{01}^2 \left(-\left(e^{\left(\frac{1.2-0.1}{\sqrt{2}\sqrt{1.2}}\right)^2} \operatorname{erfc}\left(\frac{1.2-0.1}{\sqrt{2}\sqrt{1.2}}\right) - e^{\left(\frac{2\cdot 1.2-0.1}{\sqrt{2}\sqrt{1.2}}\right)^2} \operatorname{erfc}\left(\frac{2 \cdot 1.2-0.1}{\sqrt{2}\sqrt{1.2}}\right)\right)\right) - \\
&0.1 e^{\frac{0.1^2}{2\cdot 1.2}} \left(2 - \operatorname{erfc}\left(\frac{0.1}{\sqrt{2}\sqrt{1.2}}\right)\right) + \sqrt{1.2}\sqrt{\frac{2}{\pi}} \; = \; 0.212234 \, .
\end{aligned}$$

Therefore the term in brackets of Eq. (25) is larger than zero. Thus, $\frac{\partial}{\partial \mu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ has the sign of ω. Since $\tilde{\xi}$ is a function in $\mu\omega$ (these variables only appear as this product), we have for $x = \mu\omega$

$$\frac{\partial}{\partial \nu}\tilde{\xi} \; = \; \frac{\partial}{\partial x}\tilde{\xi} \, \frac{\partial x}{\partial \mu} \; = \; \frac{\partial}{\partial x}\tilde{\xi} \, \omega \tag{36}$$

and

$$\frac{\partial}{\partial \omega}\tilde{\xi} \; = \; \frac{\partial}{\partial x}\tilde{\xi} \, \frac{\partial x}{\partial \omega} \; = \; \frac{\partial}{\partial x}\tilde{\xi} \, \mu \, . \tag{37}$$

$$\frac{\partial}{\partial \omega}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \; = \; \frac{\mu}{\omega} \, \frac{\partial}{\partial \mu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \, . \tag{38}$$

Since $\frac{\partial}{\partial \mu}\tilde{\xi}$ has the sign of ω, $\frac{\partial}{\partial \mu}\tilde{\xi}$ has the sign of μ. Therefore

$$\frac{\partial}{\partial \omega} g(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \; = \; \frac{\partial}{\partial \omega}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha 01) \tag{39}$$

has the sign of μ.

We now divide the μ-domain into $-1 \leqslant \mu \leqslant 0$ and $0 \leqslant \mu \leqslant 1$. Analogously we divide the ω-domain into $-0.1 \leqslant \omega \leqslant 0$ and $0 \leqslant \omega \leqslant 0.1$. In this domains g is strictly monotonically.

For all domains g is strictly monotonically decreasing in ν and strictly monotonically increasing in τ. Note that we now consider the range $3 \leqslant \nu \leqslant 16$. For the maximal value of g we set $\nu = 3$ (we set it to 3!) and $\tau = 1.25$.

We consider now all combination of these domains:

- $-1 \leqslant \mu \leqslant 0$ and $-0.1 \leqslant \omega \leqslant 0$:

  g is decreasing in μ and decreasing in ω. We set $\mu = -1$ and $\omega = -0.1$.

  $$g(-1, -0.1, 3, 1.25, \lambda_{01}, \alpha_{01}) \; = \; -0.0180173 \, . \tag{40}$$

- $-1 \leqslant \mu \leqslant 0$ and $0 \leqslant \omega \leqslant 0.1$:

  g is increasing in μ and decreasing in ω. We set $\mu = 0$ and $\omega = 0$.

  $$g(0, 0, 3, 1.25, \lambda_{01}, \alpha_{01}) \; = \; -0.148532 \, . \tag{41}$$

- $0 \leqslant \mu \leqslant 1$ and $-0.1 \leqslant \omega \leqslant 0$:

  g is decreasing in μ and increasing in ω. We set $\mu = 0$ and $\omega = 0$.

  $$g(0, 0, 3, 1.25, \lambda_{01}, \alpha_{01}) \; = \; -0.148532 \, . \tag{42}$$

- $0 \leqslant \mu \leqslant 1$ and $0 \leqslant \omega \leqslant 0.1$:

  g is increasing in μ and increasing in ω. We set $\mu = 1$ and $\omega = 0.1$.

  $$g(1, 0.1, 3, 1.25, \lambda_{01}, \alpha_{01}) \;=\; -0.0180173\;. \tag{43}$$

  Therefore the maximal value of g is -0.0180173.

□

A3.3 Proof of Theorem 3

First we recall Theorem 3:

Theorem (Increasing ν)**.** *We consider $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the two domains $\Omega_1^- = \{(\mu, \omega, \nu, \tau) \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.16, 0.8 \leqslant \tau \leqslant 1.25\}$ and $\Omega_2^- = \{(\mu, \omega, \nu, \tau) \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.24, 0.9 \leqslant \tau \leqslant 1.25\}$.*

The mapping of the variance $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (5) increases

$$\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \;>\; \nu \tag{44}$$

in both Ω_1^- and Ω_2^-. All fixed points (μ, ν) of mapping Eq. (5) and Eq. (4) ensure for $0.8 \leqslant \tau$ that $\tilde{\nu} > 0.16$ and for $0.9 \leqslant \tau$ that $\tilde{\nu} > 0.24$. Consequently, the variance mapping Eq. (5) and Eq. (4) ensures a lower bound on the variance ν.

Proof. The mean value theorem states that there exists a $t \in [0, 1]$ for which

$$\begin{aligned} &\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \;-\; \tilde{\xi}(\mu, \omega, \nu_{\min}, \tau, \lambda_{01}, \alpha_{01}) \;= \\ &\frac{\partial}{\partial \nu}\tilde{\xi}(\mu, \omega, \nu + t(\nu_{\min} - \nu), \tau, \lambda_{01}, \alpha_{01})\; (\nu - \nu_{\min})\;. \end{aligned} \tag{45}$$

Therefore

$$\begin{aligned} &\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \;=\; \tilde{\xi}(\mu, \omega, \nu_{\min}, \tau, \lambda_{01}, \alpha_{01}) \;+ \\ &\frac{\partial}{\partial \nu}\tilde{\xi}(\mu, \omega, \nu + t(\nu_{\min} - \nu), \tau, \lambda_{01}, \alpha_{01})\; (\nu - \nu_{\min})\;. \end{aligned} \tag{46}$$

Therefore we are interested to bound the derivative of the ξ-mapping Eq. (13) with respect to ν:

$$\begin{aligned} &\frac{\partial}{\partial \nu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \;= \\ &\frac{1}{2}\lambda^2 \tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha^2 \left(-\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2 e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right) - \right. \\ &\left. \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right)\;. \end{aligned} \tag{47}$$

The sub-term Eq. (308) enters the derivative Eq. (47) with a negative sign! According to Lemma 18, the minimal value of sub-term Eq. (308) is obtained by the largest largest ν, by the smallest τ, and the largest $y = \mu\omega = 0.01$. Also the positive term $\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2$ is multiplied by τ, which is minimized by using the smallest τ. Therefore we can use the smallest τ in whole formula Eq. (47) to lower bound it.

First we consider the domain $0.05 \leqslant \nu \leqslant 0.16$ and $0.8 \leqslant \tau \leqslant 1.25$. The factor consisting of the exponential in front of the brackets has its smallest value for $e^{-\frac{0.01 \cdot 0.01}{2 \cdot 0.05 \cdot 0.8}}$. Since erfc is monotonically decreasing we inserted the smallest argument via $\operatorname{erfc}\left(-\frac{0.01}{\sqrt{2}\sqrt{0.05 \cdot 0.8}}\right)$ in order to obtain the maximal negative contribution. Thus, applying Lemma 18, we obtain the lower bound on the derivative:

$$\frac{1}{2}\lambda^2 \tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha^2 \left(-\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2 e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right) - \right. \tag{48}$$

$$\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\Bigg) >$$
$$\frac{1}{2}0.8e^{-\frac{0.01\cdot 0.01}{2\cdot 0.05\cdot 0.8}}\lambda_{01}^2\left(\alpha_{01}^2\left(-\left(e^{\left(\frac{0.16\cdot 0.8+0.01}{\sqrt{2}\sqrt{0.16\cdot 0.8}}\right)^2}\operatorname{erfc}\left(\frac{0.16\cdot 0.8+0.01}{\sqrt{2}\sqrt{0.16\cdot 0.8}}\right)-\right.\right.\right.$$
$$\left.\left.\left.2e^{\left(\frac{2\cdot 0.16\cdot 0.8+0.01}{\sqrt{2}\sqrt{0.16\cdot 0.8}}\right)^2}\operatorname{erfc}\left(\frac{2\cdot 0.16\cdot 0.8+0.01}{\sqrt{2}\sqrt{0.16\cdot 0.8}}\right)\right)\right)-\operatorname{erfc}\left(-\frac{0.01}{\sqrt{2}\sqrt{0.05\cdot 0.8}}\right)+2\right)) > 0.969231\ .$$

For applying the mean value theorem, we require the smallest $\tilde{\nu}(\nu)$. We follow the proof of Lemma 8, which shows that at the minimum $y=\mu\omega$ must be maximal and $x=\nu\tau$ must be minimal. Thus, the smallest $\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})$ is $\tilde{\xi}(0.01,0.01,0.05,0.8,\lambda_{01},\alpha_{01})=0.0662727$ for $0.05\leqslant\nu$ and $0.8\leqslant\tau$.

Therefore the mean value theorem and the bound on $(\tilde{\mu})^2$ (Lemma 43) provide

$$\begin{aligned}&\tilde{\nu}=\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})-\left(\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})\right)^2 > \\ &0.0662727\ +\ 0.969231(\nu-0.05)-0.005\ =\ 0.01281115+0.969231\nu\ > \\ &0.08006969\cdot 0.16+0.969231\nu\geqslant 1.049301\nu\ >\ \nu\ .\end{aligned} \tag{49}$$

Next we consider the domain $0.05\leqslant\nu\leqslant 0.24$ and $0.9\leqslant\tau\leqslant 1.25$. The factor consisting of the exponential in front of the brackets has its smallest value for $e^{-\frac{0.01\cdot 0.01}{2\cdot 0.05\cdot 0.9}}$. Since erfc is monotonically decreasing we inserted the smallest argument via $\operatorname{erfc}\left(-\frac{0.01}{\sqrt{2}\sqrt{0.05\cdot 0.9}}\right)$ in order to obtain the maximal negative contribution.

Thus, applying Lemma 18, we obtain the lower bound on the derivative:

$$\frac{1}{2}\lambda^2\tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\left(\alpha^2\left(-\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right)-\right. \tag{50}$$
$$\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\Bigg) >$$
$$\frac{1}{2}0.9e^{-\frac{0.01\cdot 0.01}{2\cdot 0.05\cdot 0.9}}\lambda_{01}^2\left(\alpha_{01}^2\left(-\left(e^{\left(\frac{0.24\cdot 0.9+0.01}{\sqrt{2}\sqrt{0.24\cdot 0.9}}\right)^2}\operatorname{erfc}\left(\frac{0.24\cdot 0.9+0.01}{\sqrt{2}\sqrt{0.24\cdot 0.9}}\right)-\right.\right.\right.$$
$$\left.\left.\left.2e^{\left(\frac{2\cdot 0.24\cdot 0.9+0.01}{\sqrt{2}\sqrt{0.24\cdot 0.9}}\right)^2}\operatorname{erfc}\left(\frac{2\cdot 0.24\cdot 0.9+0.01}{\sqrt{2}\sqrt{0.24\cdot 0.9}}\right)\right)\right)-\operatorname{erfc}\left(-\frac{0.01}{\sqrt{2}\sqrt{0.05\cdot 0.9}}\right)+2\right)) > 0.976952\ .$$

For applying the mean value theorem, we require the smallest $\tilde{\nu}(\nu)$. We follow the proof of Lemma 8, which shows that at the minimum $y=\mu\omega$ must be maximal and $x=\nu\tau$ must be minimal. Thus, the smallest $\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})$ is $\tilde{\xi}(0.01,0.01,0.05,0.9,\lambda_{01},\alpha_{01})=0.0738404$ for $0.05\leqslant\nu$ and $0.9\leqslant\tau$. Therefore the mean value theorem and the bound on $(\tilde{\mu})^2$ (Lemma 43) gives

$$\begin{aligned}&\tilde{\nu}=\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})-\left(\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})\right)^2 > \\ &0.0738404\ +\ 0.976952(\nu-0.05)-0.005\ =\ 0.0199928+0.976952\nu\ > \\ &0.08330333\cdot 0.24+0.976952\nu\geqslant 1.060255\nu\ >\ \nu\ .\end{aligned} \tag{51}$$

□

A3.4 Lemmata and Other Tools Required for the Proofs

A3.4.1 Lemmata for proofing Theorem 1 (part 1): Jacobian norm smaller than one

In this section, we show that the largest singular value of the Jacobian of the mapping g is smaller than one. Therefore, g is a contraction mapping. This is even true in a larger domain than the original Ω. We do not need to restrict $\tau\in[0.95,1.1]$, but we can extend to $\tau\in[0.8,1.25]$. The range of the other variables is unchanged such that we consider the following domain throughout this section: $\mu\in[-0.1,0.1]$, $\omega\in[-0.1,0.1]$, $\nu\in[0.8,1.5]$, and $\tau\in[0.8,1.25]$.

Jacobian of the mapping. In the following, we denote two Jacobians: (1) the Jacobian $\mathcal{J}$ of the mapping $h : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\xi})$, and (2) the Jacobian $\mathcal{H}$ of the mapping $g : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\nu})$ because the influence of $\tilde{\mu}$ on $\tilde{\nu}$ is small, and many properties of the system can already be seen on $\mathcal{J}$.

$$\mathcal{J} = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{21} & \mathcal{J}_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \mu}\tilde{\mu} & \frac{\partial}{\partial \nu}\tilde{\mu} \\ \frac{\partial}{\partial \mu}\tilde{\xi} & \frac{\partial}{\partial \nu}\tilde{\xi} \end{pmatrix} \tag{52}$$

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{21} & \mathcal{H}_{22} \end{pmatrix} = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11} & \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12} \end{pmatrix} \tag{53}$$

The definition of the entries of the Jacobian $\mathcal{J}$ is:

$$\mathcal{J}_{11}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{\partial}{\partial \mu}\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \tag{54}$$
$$\frac{1}{2}\lambda\omega\left(\alpha e^{\mu\omega + \frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right)$$

$$\mathcal{J}_{12}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{\partial}{\partial \nu}\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \tag{55}$$
$$\frac{1}{4}\lambda\tau\left(\alpha e^{\mu\omega + \frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - (\alpha - 1)\sqrt{\frac{2}{\pi\nu\tau}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)$$

$$\mathcal{J}_{21}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{\partial}{\partial \mu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \tag{56}$$
$$\lambda^2\omega\left(\alpha^2\left(-e^{\mu\omega + \frac{\nu\tau}{2}}\right) \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \right.$$
$$\left.\alpha^2 e^{2\mu\omega + 2\nu\tau} \operatorname{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \mu\omega\left(2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)$$

$$\mathcal{J}_{22}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{\partial}{\partial \nu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \tag{57}$$
$$\frac{1}{2}\lambda^2\tau\left(\alpha^2\left(-e^{\mu\omega + \frac{\nu\tau}{2}}\right) \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \right.$$
$$\left.2\alpha^2 e^{2\mu\omega + 2\nu\tau} \operatorname{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right)$$

Proof sketch: Bounding the largest singular value of the Jacobian. If the largest singular value of the Jacobian is smaller than 1, then the spectral norm of the Jacobian is smaller than 1. Then the mapping Eq. (4) and Eq. (5) of the mean and variance to the mean and variance in the next layer is contracting.

We show that the largest singular value is smaller than 1 by evaluating the function $S(\mu, \omega, \nu, \tau, \lambda, \alpha)$ on a grid. Then we use the Mean Value Theorem to bound the deviation of the function S between grid points. Toward this end we have to bound the gradient of S with respect to (μ, ω, ν, τ). If all function values plus gradient times the deltas (differences between grid points and evaluated points) is still smaller than 1, then we have proofed that the function is below 1.

The singular values of the 2×2 matrix

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{58}$$

are

$$s_1 = \frac{1}{2}\left(\sqrt{(a_{11} + a_{22})^2 + (a_{21} - a_{12})^2} + \sqrt{(a_{11} - a_{22})^2 + (a_{12} + a_{21})^2}\right) \tag{59}$$

$$s_2 = \frac{1}{2}\left(\sqrt{(a_{11} + a_{22})^2 + (a_{21} - a_{12})^2} - \sqrt{(a_{11} - a_{22})^2 + (a_{12} + a_{21})^2}\right). \tag{60}$$

We used an explicit formula for the singular values [4]. We now set $\mathcal{H}_{11} = a_{11}, \mathcal{H}_{12} = a_{12}, \mathcal{H}_{21} = a_{21}, \mathcal{H}_{22} = a_{22}$ to obtain a formula for the largest singular value of the Jacobian depending on $(\mu, \omega, \nu, \tau, \lambda, \alpha)$. The formula for the largest singular value for the Jacobian is:

$$
\begin{aligned}
S(\mu, \omega, \nu, \tau, \lambda, \alpha) \; &= \left(\sqrt{(\mathcal{H}_{11} + \mathcal{H}_{22})^2 + (\mathcal{H}_{21} - \mathcal{H}_{12})^2} \; + \; \sqrt{(\mathcal{H}_{11} - \mathcal{H}_{22})^2 + (\mathcal{H}_{12} + \mathcal{H}_{21})^2} \right) = \qquad (61) \\
&= \frac{1}{2} \left(\sqrt{(\mathcal{J}_{11} + \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12})^2 + (\mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11} - \mathcal{J}_{12})^2} + \right. \\
&\left. \sqrt{(\mathcal{J}_{11} - \mathcal{J}_{22} + 2\tilde{\mu}\mathcal{J}_{12})^2 + (\mathcal{J}_{12} + \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11})^2} \right),
\end{aligned}
$$

where $\mathcal{J}$ are defined in Eq. (54) and we left out the dependencies on $(\mu, \omega, \nu, \tau, \lambda, \alpha)$ in order to keep the notation uncluttered, e.g. we wrote $\mathcal{J}_{11}$ instead of $\mathcal{J}_{11}(\mu, \omega, \nu, \tau, \lambda, \alpha)$.

Bounds on the derivatives of the Jacobian entries. In order to bound the gradient of the singular value, we have to bound the derivatives of the Jacobian entries $\mathcal{J}_{11}(\mu, \omega, \nu, \tau, \lambda, \alpha)$, $\mathcal{J}_{12}(\mu, \omega, \nu, \tau, \lambda, \alpha)$, $\mathcal{J}_{21}(\mu, \omega, \nu, \tau, \lambda, \alpha)$, and $\mathcal{J}_{22}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ with respect to μ, ω, ν, and τ. The values λ and α are fixed to λ_{01} and α_{01}. The 16 derivatives of the 4 Jacobian entries with respect to the 4 variables are:

$$
\begin{aligned}
\frac{\partial \mathcal{J}_{11}}{\partial \mu} &= \frac{1}{2} \lambda \omega^2 e^{-\frac{\mu^2 \omega^2}{2 \nu \tau}} \left(\alpha e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) - \frac{\sqrt{\frac{2}{\pi}}(\alpha - 1)}{\sqrt{\nu\tau}} \right) \qquad (62) \\
\frac{\partial \mathcal{J}_{11}}{\partial \omega} &= \frac{1}{2} \lambda \left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\frac{\sqrt{\frac{2}{\pi}}(\alpha - 1)\mu\omega}{\sqrt{\nu\tau}} - \alpha(\mu\omega + 1) e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) \right) - \right. \\
&\left. \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) + 2 \right) \\
\frac{\partial \mathcal{J}_{11}}{\partial \nu} &= \frac{1}{4} \lambda \tau \omega e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \sqrt{\frac{2}{\pi}} \left(\frac{(\alpha - 1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}} \right) \right) \\
\frac{\partial \mathcal{J}_{11}}{\partial \tau} &= \frac{1}{4} \lambda \nu \omega e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \sqrt{\frac{2}{\pi}} \left(\frac{(\alpha - 1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}} \right) \right) \\
\frac{\partial \mathcal{J}_{12}}{\partial \mu} &= \frac{\partial \mathcal{J}_{11}}{\partial \nu} \\
\frac{\partial \mathcal{J}_{12}}{\partial \omega} &= \frac{1}{4} \lambda \mu \tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \sqrt{\frac{2}{\pi}} \left(\frac{(\alpha - 1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}} \right) \right) \\
\frac{\partial \mathcal{J}_{12}}{\partial \nu} &= \frac{1}{8} \lambda e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\alpha \tau^2 e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \right. \\
&\left. \sqrt{\frac{2}{\pi}} \left(\frac{(-1)(\alpha - 1)\mu^2\omega^2}{\nu^{5/2}\sqrt{\tau}} + \frac{\sqrt{\tau}(\alpha + \alpha\mu\omega - 1)}{\nu^{3/2}} - \frac{\alpha\tau^{3/2}}{\sqrt{\nu}} \right) \right) \\
\frac{\partial \mathcal{J}_{12}}{\partial \tau} &= \frac{1}{8} \lambda e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(2\alpha e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \alpha\nu\tau e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \right. \\
&\left. \sqrt{\frac{2}{\pi}} \left(\frac{(-1)(\alpha - 1)\mu^2\omega^2}{(\nu\tau)^{3/2}} + \frac{-\alpha + \alpha\mu\omega + 1}{\sqrt{\nu\tau}} - \alpha\sqrt{\nu\tau} \right) \right) \\
\frac{\partial \mathcal{J}_{21}}{\partial \mu} &= \lambda^2 \omega^2 \left(\alpha^2 \left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \right) e^{\frac{(\mu\omega + \nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad 2\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\Bigg) \\
\frac{\partial \mathcal{J}_{21}}{\partial \omega} &= \lambda^2 \Bigg(\alpha^2(\mu\omega+1)\left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right) e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad \alpha^2(2\mu\omega+1) e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 2\mu\omega\left(2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\Bigg) \\
\frac{\partial \mathcal{J}_{21}}{\partial \nu} &= \frac{1}{2}\lambda^2 \tau\omega e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \Bigg(\alpha^2\left(-e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\right) \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 4\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha^2-1\right)}{\sqrt{\nu\tau}}\Bigg) \\
\frac{\partial \mathcal{J}_{21}}{\partial \tau} &= \frac{1}{2}\lambda^2 \nu\omega e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \Bigg(\alpha^2\left(-e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\right) \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 4\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha^2-1\right)}{\sqrt{\nu\tau}}\Bigg) \\
\frac{\partial \mathcal{J}_{22}}{\partial \mu} &= \frac{\partial \mathcal{J}_{21}}{\partial \nu} \\
\frac{\partial \mathcal{J}_{22}}{\partial \omega} &= \frac{1}{2}\lambda^2 \mu\tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \Bigg(\alpha^2\left(-e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\right) \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 4\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha^2-1\right)}{\sqrt{\nu\tau}}\Bigg) \\
\frac{\partial \mathcal{J}_{22}}{\partial \nu} &= \frac{1}{4}\lambda^2 \tau^2 e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \Bigg(\alpha^2\left(-e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\right) \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 8\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha^2-1\right)\mu\omega}{(\nu\tau)^{3/2}} - \frac{3\alpha^2}{\sqrt{\nu\tau}}\right)\Bigg) \\
\frac{\partial \mathcal{J}_{22}}{\partial \tau} &= \frac{1}{4}\lambda^2 \Bigg(-2\alpha^2 e^{-\frac{\mu^2\omega^2}{2\nu\tau}} e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \\
&\quad \alpha^2\nu\tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}} e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + 4\alpha^2 e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \\
&\quad 8\alpha^2\nu\tau e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\left(2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \\
&\quad \sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \left(\frac{\left(\alpha^2-1\right)\mu\omega}{\sqrt{\nu\tau}} - 3\alpha^2\sqrt{\nu\tau}\right)\Bigg)
\end{aligned}
$$

Lemma 5 (Bounds on the Derivatives)**.** *The following bounds on the absolute values of the derivatives of the Jacobian entries $\mathcal{J}_{11}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, $\mathcal{J}_{12}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, $\mathcal{J}_{21}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, and $\mathcal{J}_{22}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ with respect to μ, ω, ν, and τ hold:*

$$
\begin{aligned}
\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right| &< 0.0031049101995398316 \qquad (63) \\
\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| &< 1.055872374194189
\end{aligned}
$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| < 0.03749149348255419$$

$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| < 0.21232788238624354$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| < 0.2124377655377270$$

$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \mu}\right| < 0.02220441024325437$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \omega}\right| < 1.146955401845684$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \nu}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \tau}\right| < 0.17980135762932363$$

$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \mu}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \omega}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \nu}\right| < 1.805740052651535$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \tau}\right| < 2.396685907216327$$

Proof. See proof 39. □

Bounds on the entries of the Jacobian.

Lemma 6 (Bound on J11). *The absolute value of the function* $\mathcal{J}_{11} = \frac{1}{2}\lambda\omega\left(\alpha e^{\mu\omega+\frac{\nu\tau}{2}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right)$ *is bounded by* $|\mathcal{J}_{11}| \leqslant 0.104497$ *in the domain* $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$, $0.8 \leqslant \nu \leqslant 1.5$, *and* $0.8 \leqslant \tau \leqslant 1.25$ *for* $\alpha = \alpha_{01}$ *and* $\lambda = \lambda_{01}$.

Proof.

$$\begin{aligned}|\mathcal{J}_{11}| &= \left|\frac{1}{2}\lambda\omega\left(\alpha e^{\mu\omega+\frac{\nu\tau}{2}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right| \\ &\leqslant |\frac{1}{2}||\lambda||\omega|\left(|\alpha|0.587622 + 1.00584\right) \leqslant 0.104497,\end{aligned}$$

(64)

where we used that (a) J_{11} is strictly monotonically increasing in $\mu\omega$ and $|2 - \operatorname{erfc}\left(\frac{0.01}{\sqrt{2}\sqrt{\nu\tau}}\right)| \leqslant 1.00584$ and (b) Lemma 47 that $|e^{\mu\omega+\frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)| \leqslant e^{0.01+\frac{0.64}{2}} \operatorname{erfc}\left(\frac{0.01+0.64}{\sqrt{2}\sqrt{0.64}}\right) = 0.587622$ $\square$

Lemma 7 (Bound on J12)**.** *The absolute value of the function* $\mathcal{J}_{12} = \frac{1}{4}\lambda\tau\left(\alpha e^{\mu\omega+\frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - (\alpha-1)\sqrt{\frac{2}{\pi\nu\tau}}e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)$ *is bounded by* $|\mathcal{J}_{12}| \leqslant 0.194145$ *in the domain* $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$, $0.8 \leqslant \nu \leqslant 1.5$, *and* $0.8 \leqslant \tau \leqslant 1.25$ *for* $\alpha = \alpha_{01}$ *and* $\lambda = \lambda_{01}$.

Proof.

$$
\begin{aligned}
&|J_{12}| \leqslant \frac{1}{4}|\lambda||\tau| \left|\left(\alpha e^{\mu\omega+\frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - (\alpha-1)\sqrt{\frac{2}{\pi\nu\tau}}e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)\right| \leqslant \\
&\frac{1}{4}|\lambda||\tau|\,|0.983247 - 0.392294| \leqslant \\
&0.194035
\end{aligned} \tag{65}
$$

For the first term we have $0.434947 \leqslant e^{\mu\omega+\frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) \leqslant 0.587622$ after Lemma 47 and for the second term $0.582677 \leqslant \sqrt{\frac{2}{\pi\nu\tau}}e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \leqslant 0.997356$, which can easily be seen by maximizing or minimizing the arguments of the exponential or the square root function. The first term scaled by α is $0.727780 \leqslant \alpha e^{\mu\omega+\frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) \leqslant 0.983247$ and the second term scaled by $\alpha - 1$ is $0.392294 \leqslant (\alpha-1)\sqrt{\frac{2}{\pi\nu\tau}}e^{-\frac{\mu^2\omega^2}{2\nu\tau}} \leqslant 0.671484$. Therefore, the absolute difference between these terms is at most $0.983247 - 0.392294$ leading to the derived bound.

$\square$

Bounds on mean, variance and second moment. For deriving bounds on $\tilde{\mu}$, $\tilde{\xi}$, and $\tilde{\nu}$, we need the following lemma.

Lemma 8 (Derivatives of the Mapping)**.** *We assume* $\alpha = \alpha_{01}$ *and* $\lambda = \lambda_{01}$. *We restrict the range of the variables to the domain* $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, *and* $\tau \in [0.8, 1.25]$.

The derivative $\frac{\partial}{\partial\mu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ *has the sign of* ω.

The derivative $\frac{\partial}{\partial\nu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ *is positive.*

The derivative $\frac{\partial}{\partial\mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ *has the sign of* ω.

The derivative $\frac{\partial}{\partial\nu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ *is positive.*

Proof. See 40. $\square$

Lemma 9 (Bounds on mean, variance and second moment)**.** *The expressions* $\tilde{\mu}$, $\tilde{\xi}$, *and* $\tilde{\nu}$ *for* $\alpha = \alpha_{01}$ *and* $\lambda = \lambda_{01}$ *are bounded by* $-0.041160 < \tilde{\mu} < 0.087653$, $0.703257 < \tilde{\xi} < 1.643705$ *and* $0.695574 < \tilde{\nu} < 1.636023$ *in the domain* $\mu \in [-0.1, 0.1]$, $\nu \in [0.8, 15]$, $\omega \in [-0.1, 0.1]$, $\tau \in [0.8, 1.25]$.

Proof. We use Lemma 8 which states that with given sign the derivatives of the mapping Eq. (4) and Eq. (5) with respect to ν and μ are either positive or have the sign of ω. Therefore with given sign of ω the mappings are strict monotonic and the their maxima and minima are found at the borders. The minimum of $\tilde{\mu}$ is obtained at $\mu\omega = -0.01$ and its maximum at $\mu\omega = 0.01$ and σ and τ at minimal or maximal values, respectively. It follows that

$$
-0.041160 < \tilde{\mu}(-0.1, 0.1, 0.8, 0.8, \lambda_{01}, \alpha_{01}) \leqslant \tilde{\mu} \leqslant \tilde{\mu}(0.1, 0.1, 1.5, 1.25, \lambda_{01}, \alpha_{01}) < 0.087653. \tag{66}
$$

Similarly, the maximum and minimum of $\tilde{\xi}$ is obtained at the values mentioned above:

$$0.703257 < \tilde{\xi}(-0.1, 0.1, 0.8, 0.8, \lambda_{01}, \alpha_{01}) \leqslant \tilde{\xi} \leqslant \tilde{\xi}(0.1, 0.1, 1.5, 1.25, \lambda_{01}, \alpha_{01}) < 1.643705. \tag{67}$$

Hence we obtain the following bounds on $\tilde{\nu}$:

$$\begin{aligned}
0.703257 - \tilde{\mu}^2 < \tilde{\xi} - \tilde{\mu}^2 < 1.643705 - \tilde{\mu}^2 \\
0.703257 - 0.007683 < \tilde{\nu} < 1.643705 - 0.007682 \\
0.695574 < \tilde{\nu} < 1.636023.
\end{aligned} \tag{68}$$

□

Upper Bounds on the Largest Singular Value of the Jacobian.

Lemma 10 (Upper Bounds on Absolute Derivatives of Largest Singular Value)**.** *We set $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ and restrict the range of the variables to $\mu \in [\mu_{\min}, \mu_{\max}] = [-0.1, 0.1]$, $\omega \in [\omega_{\min}, \omega_{\max}] = [-0.1, 0.1]$, $\nu \in [\nu_{\min}, \nu_{\max}] = [0.8, 1.5]$, and $\tau \in [\tau_{\min}, \tau_{\max}] = [0.8, 1.25]$.*

The absolute values of derivatives of the largest singular value $S(\mu, \omega, \nu, \tau, \lambda, \alpha)$ given in Eq. (61) with respect to (μ, ω, ν, τ) are bounded as follows:

$$\left| \frac{\partial S}{\partial \mu} \right| < 0.32112 , \tag{69}$$

$$\left| \frac{\partial S}{\partial \omega} \right| < 2.63690 , \tag{70}$$

$$\left| \frac{\partial S}{\partial \nu} \right| < 2.28242 , \tag{71}$$

$$\left| \frac{\partial S}{\partial \tau} \right| < 2.98610 . \tag{72}$$

Proof. The Jacobian of our mapping Eq. (4) and Eq. (5) is defined as

$$\boldsymbol{H} = \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{21} & \mathcal{H}_{22} \end{pmatrix} = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11} & \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12} \end{pmatrix} \tag{73}$$

and has the largest singular value

$$S(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{1}{2}\left(\sqrt{(\mathcal{H}_{11} - \mathcal{H}_{22})^2 + (\mathcal{H}_{12} + \mathcal{H}_{21})^2} + \sqrt{(\mathcal{H}_{11} + \mathcal{H}_{22})^2 + (\mathcal{H}_{12} - \mathcal{H}_{21})^2}\right), \tag{74}$$

according to the formula of Blinn [4].

We obtain

$$\left| \frac{\partial S}{\partial \mathcal{H}_{11}} \right| = \left| \frac{1}{2}\left(\frac{\mathcal{H}_{11} - \mathcal{H}_{22}}{\sqrt{(\mathcal{H}_{11} - \mathcal{H}_{22})^2 + (\mathcal{H}_{12} + \mathcal{H}_{21})^2}} + \frac{\mathcal{H}_{11} + \mathcal{H}_{22}}{\sqrt{(\mathcal{H}_{11} + \mathcal{H}_{22})^2 + (\mathcal{H}_{21} - \mathcal{H}_{12})^2}} \right) \right| < \tag{75}$$

$$\frac{1}{2}\left(\left| \frac{1}{\sqrt{\frac{(\mathcal{H}_{12}+\mathcal{H}_{21})^2}{(\mathcal{H}_{11}-\mathcal{H}_{22})^2} + 1}} \right| + \left| \frac{1}{\sqrt{\frac{(\mathcal{H}_{21}-\mathcal{H}_{12})^2}{(\mathcal{H}_{11}+\mathcal{H}_{22})^2} + 1}} \right| \right) < \frac{1+1}{2} = 1$$

and analogously

$$\left| \frac{\partial S}{\partial \mathcal{H}_{12}} \right| = \left| \frac{1}{2}\left(\frac{\mathcal{H}_{12} + \mathcal{H}_{21}}{\sqrt{(\mathcal{H}_{11} - \mathcal{H}_{22})^2 + (\mathcal{H}_{12} + \mathcal{H}_{21})^2}} - \frac{\mathcal{H}_{21} - \mathcal{H}_{12}}{\sqrt{(\mathcal{H}_{11} + \mathcal{H}_{22})^2 + (\mathcal{H}_{21} - \mathcal{H}_{12})^2}} \right) \right| < 1 \tag{76}$$

and

$$\left|\frac{\partial S}{\partial \mathcal{H}_{21}}\right| = \left|\frac{1}{2}\left(\frac{\mathcal{H}_{21}-\mathcal{H}_{12}}{\sqrt{(\mathcal{H}_{11}+\mathcal{H}_{22})^2+(\mathcal{H}_{21}-\mathcal{H}_{12})^2}} + \frac{\mathcal{H}_{12}+\mathcal{H}_{21}}{\sqrt{(\mathcal{H}_{11}-\mathcal{H}_{22})^2+(\mathcal{H}_{12}+\mathcal{H}_{21})^2}}\right)\right| < 1 \tag{77}$$

and

$$\left|\frac{\partial S}{\partial \mathcal{H}_{22}}\right| = \left|\frac{1}{2}\left(\frac{\mathcal{H}_{11}+\mathcal{H}_{22}}{\sqrt{(\mathcal{H}_{11}+\mathcal{H}_{22})^2+(\mathcal{H}_{21}-\mathcal{H}_{12})^2}} - \frac{\mathcal{H}_{11}-\mathcal{H}_{22}}{\sqrt{(\mathcal{H}_{11}-\mathcal{H}_{22})^2+(\mathcal{H}_{12}+\mathcal{H}_{21})^2}}\right)\right| < 1\,. \tag{78}$$

We have

$$\frac{\partial S}{\partial \mu} = \frac{\partial S}{\partial \mathcal{H}_{11}}\frac{\partial \mathcal{H}_{11}}{\partial \mu} + \frac{\partial S}{\partial \mathcal{H}_{12}}\frac{\partial \mathcal{H}_{12}}{\partial \mu} + \frac{\partial S}{\partial \mathcal{H}_{21}}\frac{\partial \mathcal{H}_{21}}{\partial \mu} + \frac{\partial S}{\partial \mathcal{H}_{22}}\frac{\partial \mathcal{H}_{22}}{\partial \mu} \tag{79}$$

$$\frac{\partial S}{\partial \omega} = \frac{\partial S}{\partial \mathcal{H}_{11}}\frac{\partial \mathcal{H}_{11}}{\partial \omega} + \frac{\partial S}{\partial \mathcal{H}_{12}}\frac{\partial \mathcal{H}_{12}}{\partial \omega} + \frac{\partial S}{\partial \mathcal{H}_{21}}\frac{\partial \mathcal{H}_{21}}{\partial \omega} + \frac{\partial S}{\partial \mathcal{H}_{22}}\frac{\partial \mathcal{H}_{22}}{\partial \omega} \tag{80}$$

$$\frac{\partial S}{\partial \nu} = \frac{\partial S}{\partial \mathcal{H}_{11}}\frac{\partial \mathcal{H}_{11}}{\partial \nu} + \frac{\partial S}{\partial \mathcal{H}_{12}}\frac{\partial \mathcal{H}_{12}}{\partial \nu} + \frac{\partial S}{\partial \mathcal{H}_{21}}\frac{\partial \mathcal{H}_{21}}{\partial \nu} + \frac{\partial S}{\partial \mathcal{H}_{22}}\frac{\partial \mathcal{H}_{22}}{\partial \nu} \tag{81}$$

$$\frac{\partial S}{\partial \tau} = \frac{\partial S}{\partial \mathcal{H}_{11}}\frac{\partial \mathcal{H}_{11}}{\partial \tau} + \frac{\partial S}{\partial \mathcal{H}_{12}}\frac{\partial \mathcal{H}_{12}}{\partial \tau} + \frac{\partial S}{\partial \mathcal{H}_{21}}\frac{\partial \mathcal{H}_{21}}{\partial \tau} + \frac{\partial S}{\partial \mathcal{H}_{22}}\frac{\partial \mathcal{H}_{22}}{\partial \tau} \tag{82}$$

$$\tag{83}$$

from which follows using the bounds from Lemma 5:

Derivative of the singular value w.r.t. μ:

$$\begin{aligned}
&\left|\frac{\partial S}{\partial \mu}\right| \leqslant \qquad (84)\\
&\left|\frac{\partial S}{\partial \mathcal{H}_{11}}\right|\left|\frac{\partial \mathcal{H}_{11}}{\partial \mu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{12}}\right|\left|\frac{\partial \mathcal{H}_{12}}{\partial \mu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{21}}\right|\left|\frac{\partial \mathcal{H}_{21}}{\partial \mu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{22}}\right|\left|\frac{\partial \mathcal{H}_{22}}{\partial \mu}\right| \leqslant\\
&\left|\frac{\partial \mathcal{H}_{11}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{H}_{12}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{H}_{21}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{H}_{22}}{\partial \mu}\right| \leqslant\\
&\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12}}{\partial \mu}\right| \leqslant\\
&\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{21}}{\partial \mu}\right| + \left|\frac{\partial \mathcal{J}_{22}}{\partial \mu}\right| + 2\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right||\tilde{\mu}| + 2\left|\mathcal{J}_{11}\right|^2 + 2\left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right||\tilde{\mu}| + 2\left|\mathcal{J}_{12}\right|\left|\mathcal{J}_{11}\right| \leqslant\\
&0.0031049101995398316 + 0.031242911235461816 + 0.02220441024325437 + 0.14983446469110305 +\\
&2 \cdot 0.104497 \cdot 0.087653 + 2 \cdot 0.104497^2 +\\
&2 \cdot 0.194035 \cdot 0.087653 + 2 \cdot 0.104497 \cdot 0.194035 < 0.32112,
\end{aligned}$$

where we used the results from the lemmata 5, 6, 7, and 9.

Derivative of the singular value w.r.t. ω:

$$\begin{aligned}
&\left|\frac{\partial S}{\partial \omega}\right| \leqslant \qquad (85)\\
&\left|\frac{\partial S}{\partial \mathcal{H}_{11}}\right|\left|\frac{\partial \mathcal{H}_{11}}{\partial \omega}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{12}}\right|\left|\frac{\partial \mathcal{H}_{12}}{\partial \omega}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{21}}\right|\left|\frac{\partial \mathcal{H}_{21}}{\partial \omega}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{22}}\right|\left|\frac{\partial \mathcal{H}_{22}}{\partial \omega}\right| \leqslant\\
&\left|\frac{\partial \mathcal{H}_{11}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{H}_{12}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{H}_{21}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{H}_{22}}{\partial \omega}\right| \leqslant\\
&\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12}}{\partial \omega}\right| \leqslant
\end{aligned}$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{21}}{\partial \omega}\right| + \left|\frac{\partial \mathcal{J}_{22}}{\partial \omega}\right| + 2\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{11}| \left|\frac{\partial \tilde{\mu}}{\partial \omega}\right| +$$

$$2\left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{12}| \left|\frac{\partial \tilde{\mu}}{\partial \omega}\right| \leqslant \tag{86}$$

$$2.38392 + 2 \cdot 1.055872374194189 \cdot 0.087653 + 2 \cdot 0.104497^2 + 2 \cdot 0.031242911235461816 \cdot 0.087653$$
$$+ 2 \cdot 0.194035 \cdot 0.104497 \ < \ 2.63690 \ ,$$

where we used the results from the lemmata 5, 6, 7, and 9 and that $\tilde{\mu}$ is symmetric for μ, ω.

Derivative of the singular value w.r.t. ν:

$$\left|\frac{\partial S}{\partial \nu}\right| \leqslant \tag{87}$$

$$\left|\frac{\partial S}{\partial \mathcal{H}_{11}}\right| \left|\frac{\partial \mathcal{H}_{11}}{\partial \nu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{12}}\right| \left|\frac{\partial \mathcal{H}_{12}}{\partial \nu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{21}}\right| \left|\frac{\partial \mathcal{H}_{21}}{\partial \nu}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{22}}\right| \left|\frac{\partial \mathcal{H}_{22}}{\partial \nu}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{H}_{11}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{H}_{12}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{H}_{21}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{H}_{22}}{\partial \nu}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12}}{\partial \nu}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{21}}{\partial \nu}\right| + \left|\frac{\partial \mathcal{J}_{22}}{\partial \nu}\right| + 2\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{11}|\,|\mathcal{J}_{12}| + 2\left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{12}|^2 \leqslant$$

$$2.19916 + 2 \cdot 0.031242911235461816 \cdot 0.087653 + 2 \cdot 0.104497 \cdot 0.194035 +$$
$$2 \cdot 0.21232788238624354 \cdot 0.087653 + 2 \cdot 0.194035^2 \ < \ 2.28242 \ ,$$

where we used the results from the lemmata 5, 6, 7, and 9.

Derivative of the singular value w.r.t. τ:

$$\left|\frac{\partial S}{\partial \tau}\right| \leqslant \tag{88}$$

$$\left|\frac{\partial S}{\partial \mathcal{H}_{11}}\right| \left|\frac{\partial \mathcal{H}_{11}}{\partial \tau}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{12}}\right| \left|\frac{\partial \mathcal{H}_{12}}{\partial \tau}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{21}}\right| \left|\frac{\partial \mathcal{H}_{21}}{\partial \tau}\right| + \left|\frac{\partial S}{\partial \mathcal{H}_{22}}\right| \left|\frac{\partial \mathcal{H}_{22}}{\partial \tau}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{H}_{11}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{H}_{12}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{H}_{21}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{H}_{22}}{\partial \tau}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{21} - 2\tilde{\mu}\mathcal{J}_{11}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{22} - 2\tilde{\mu}\mathcal{J}_{12}}{\partial \tau}\right| \leqslant$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{21}}{\partial \tau}\right| + \left|\frac{\partial \mathcal{J}_{22}}{\partial \tau}\right| + 2\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{11}| \left|\frac{\partial \tilde{\mu}}{\partial \tau}\right| +$$

$$2\left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| |\tilde{\mu}| + 2\,|\mathcal{J}_{12}| \left|\frac{\partial \tilde{\mu}}{\partial \tau}\right| \leqslant \tag{89}$$

$$2.82643 + 2 \cdot 0.03749149348255419 \cdot 0.087653 + 2 \cdot 0.104497 \cdot 0.194035 +$$
$$2 \cdot 0.2124377655377270 \cdot 0.087653 + 2 \cdot 0.194035^2 \ < \ 2.98610 \ ,$$

where we used the results from the lemmata 5, 6, 7, and 9 and that $\tilde{\mu}$ is symmetric for ν, τ.

□

Lemma 11 (Mean Value Theorem Bound on Deviation from Largest Singular Value)**.** *We set $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ and restrict the range of the variables to $\mu \in [\mu_{\min}, \mu_{\max}] = [-0.1, 0.1]$, $\omega \in [\omega_{\min}, \omega_{\max}] = [-0.1, 0.1]$, $\nu \in [\nu_{\min}, \nu_{\max}] = [0.8, 1.5]$, and $\tau \in [\tau_{\min}, \tau_{\max}] = [0.8, 1.25]$.*

The distance of the singular value at $S(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})$ and that at $S(\mu + \Delta\mu, \omega + \Delta\omega, \nu + \Delta\nu, \tau + \Delta\tau, \lambda_{01}, \alpha_{01})$ is bounded as follows:

$$|S(\mu + \Delta\mu, \omega + \Delta\omega, \nu + \Delta\nu, \tau + \Delta\tau, \lambda_{01}, \alpha_{01}) \ - \ S(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})| \ < \tag{90}$$

$$0.32112\,|\Delta\mu| + 2.63690\,|\Delta\omega| + 2.28242\,|\Delta\nu| + 2.98610\,|\Delta\tau| \ .$$

Proof. The mean value theorem states that a $t \in [0,1]$ exists for which

$$\begin{aligned}
& S(\mu+\Delta\mu,\omega+\Delta\omega,\nu+\Delta\nu,\tau+\Delta\tau,\lambda_{01},\alpha_{01}) \ - \ S(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01}) \ = \\
& \frac{\partial S}{\partial\mu}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\ \Delta\mu \ + \\
& \frac{\partial S}{\partial\omega}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\ \Delta\omega \ + \\
& \frac{\partial S}{\partial\nu}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\ \Delta\nu \ + \\
& \frac{\partial S}{\partial\tau}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\ \Delta\tau
\end{aligned} \tag{91}$$

from which immediately follows that

$$\begin{aligned}
& |S(\mu+\Delta\mu,\omega+\Delta\omega,\nu+\Delta\nu,\tau+\Delta\tau,\lambda_{01},\alpha_{01}) \ - \ S(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})| \ \leqslant \\
& \left|\frac{\partial S}{\partial\mu}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\right| \ |\Delta\mu| \ + \\
& \left|\frac{\partial S}{\partial\omega}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\right| \ |\Delta\omega| \ + \\
& \left|\frac{\partial S}{\partial\nu}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\right| \ |\Delta\nu| \ + \\
& \left|\frac{\partial S}{\partial\tau}(\mu+t\Delta\mu,\omega+t\Delta\omega,\nu+t\Delta\nu,\tau+t\Delta\tau,\lambda_{01},\alpha_{01})\right| \ |\Delta\tau| \ .
\end{aligned} \tag{92}$$

We now apply Lemma 10 which gives bounds on the derivatives, which immediately gives the statement of the lemma. □

Lemma 12 (Largest Singular Value Smaller Than One)**.** *We set $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ and restrict the range of the variables to $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.8, 1.25]$.*

The the largest singular value of the Jacobian is smaller than 1:

$$S(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01}) \ < \ 1 \ . \tag{93}$$

Therefore the mapping Eq. (4) and Eq. (5) is a contraction mapping.

Proof. We set $\Delta\mu = 0.0068097371$, $\Delta\omega = 0.0008292885$, $\Delta\nu = 0.0009580840$, and $\Delta\tau = 0.0007323095$.

According to Lemma 11 we have

$$\begin{aligned}
& |S(\mu+\Delta\mu,\omega+\Delta\omega,\nu+\Delta\nu,\tau+\Delta\tau,\lambda_{01},\alpha_{01}) \ - \ S(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})| \ < \\
& 0.32112 \cdot 0.0068097371 + 2.63690 \cdot 0.0008292885 + \\
& 2.28242 \cdot 0.0009580840 + 2.98610 \cdot 0.0007323095 \ < \ 0.008747 \ .
\end{aligned} \tag{94}$$

For a grid with grid length $\Delta\mu = 0.0068097371$, $\Delta\omega = 0.0008292885$, $\Delta\nu = 0.0009580840$, and $\Delta\tau = 0.0007323095$, we evaluated the function Eq. (61) for the largest singular value in the domain $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.8, 1.25]$. We did this using a computer. According to Subsection A3.4.5 the precision if regarding error propagation and precision of the implemented functions is larger than 10^{-13}. We performed the evaluation on different operating systems and different hardware architectures including CPUs and GPUs. In all cases the function Eq. (61) for the largest singular value of the Jacobian is bounded by 0.9912524171058772.

We obtain from Eq. (94):

$$S(\mu+\Delta\mu,\omega+\Delta\omega,\nu+\Delta\nu,\tau+\Delta\tau,\lambda_{01},\alpha_{01}) \ \leqslant \ 0.9912524171058772 \ + \ 0.008747 \ < \ 1 \ . \tag{95}$$

□

A3.4.2 Lemmata for proofing Theorem 1 (part 2): Mapping within domain

We further have to investigate whether the the mapping Eq. (4) and Eq. (5) maps into a predefined domains.

Lemma 13 (Mapping into the domain)**.** *The mapping Eq. (4) and Eq. (5) map for $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ into the domain $\mu \in [-0.03106, 0.06773]$ and $\nu \in [0.80009, 1.48617]$ with $\omega \in [-0.1, 0.1]$ and $\tau \in [0.95, 1.1]$.*

Proof. We use Lemma 8 which states that with given sign the derivatives of the mapping Eq. (4) and Eq. (5) with respect to $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$ are either positive or have the sign of ω. Therefore with given sign of ω the mappings are strict monotonic and the their maxima and minima are found at the borders. The minimum of $\tilde{\mu}$ is obtained at $\mu\omega = -0.01$ and its maximum at $\mu\omega = 0.01$ and σ and τ at their minimal and maximal values, respectively. It follows that:

$$-0.03106 < \tilde{\mu}(-0.1, 0.1, 0.8, 0.95, \lambda_{01}, \alpha_{01}) \leqslant \tilde{\mu} \leqslant \tilde{\mu}(0.1, 0.1, 1.5, 1.1, \lambda_{01}, \alpha_{01}) < 0.06773, \tag{96}$$

and that $\tilde{\mu} \in [-0.1, 0.1]$.

Similarly, the maximum and minimum of $\tilde{\xi}($ is obtained at the values mentioned above:

$$0.80467 < \tilde{\xi}(-0.1, 0.1, 0.8, 0.95, \lambda_{01}, \alpha_{01}) \leqslant \tilde{\xi} \leqslant \tilde{\xi}(0.1, 0.1, 1.5, 1.1, \lambda_{01}, \alpha_{01}) < 1.48617. \tag{97}$$

Since $|\tilde{\xi} - \tilde{\nu}| = |\tilde{\mu}^2| < 0.004597$, we can conclude that $0.80009 < \tilde{\nu} < 1.48617$ and the variance remains in $[0.8, 1.5]$. $\square$

Corollary 14. *The image $g(\Omega')$ of the mapping $g : (\mu, \nu) \mapsto (\tilde{\mu}, \tilde{\nu})$ (Eq. (8)) and the domain $\Omega' = \{(\mu, \nu) | -0.1 \leqslant \mu \leqslant 0.1, 0.8 \leqslant \mu \leqslant 1.5\}$ is a subset of Ω':*

$$g(\Omega') \subseteq \Omega', \tag{98}$$

for all $\omega \in [-0.1, 0.1]$ and $\tau \in [0.95, 1.1]$.

Proof. Directly follows from Lemma 13. $\square$

A3.4.3 Lemmata for proofing Theorem 2: The variance is contracting

Main Sub-Function. We consider the main sub-function of the derivate of second moment, $J22$ (Eq. (54)):

$$\frac{\partial}{\partial \nu} \tilde{\xi} = \frac{1}{2} \lambda^2 \tau \left(-\alpha^2 e^{\mu\omega + \frac{\nu\tau}{2}} \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\alpha^2 e^{2\mu\omega + 2\nu\tau} \operatorname{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2 \right) \tag{99}$$

that depends on $\mu\omega$ and $\nu\tau$, therefore we set $x = \nu\tau$ and $y = \mu\omega$. Algebraic reformulations provide the formula in the following form:

$$\frac{\partial}{\partial \nu} \tilde{\xi} = \frac{1}{2} \lambda^2 \tau \left(\alpha^2 \left(-e^{-\frac{y^2}{2x}} \right) \left(e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{y+x}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{y+2x}{\sqrt{2}\sqrt{x}}\right) \right) - \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{x}}\right) + 2 \right) \tag{100}$$

For $\lambda = \lambda_{01}$ and $\alpha = \alpha_{01}$, we consider the domain $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$, $1.5 \leqslant \nu \leqslant 16$, and, $0.8 \leqslant \tau \leqslant 1.25$.

For x and y we obtain: $0.8 \cdot 1.5 = 1.2 \leqslant x \leqslant 20 = 1.25 \cdot 16$ and $0.1 \cdot (-1) = -0.1 \leqslant y \leqslant 0.1 = 0.1 \cdot 1$. In the following we assume to remain within this domain.

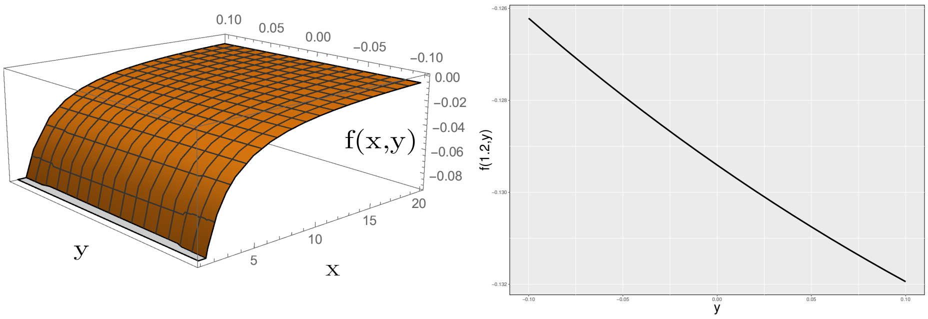

Figure A3: **Left panel:** Graphs of the main subfunction $f(x,y) = e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)$ treated in Lemma 15. The function is negative and monotonically increasing with x independent of y. **Right panel:** Graphs of the main subfunction at minimal $x = 1.2$. The graph shows that the function $f(1.2, y)$ is strictly monotonically decreasing in y.

Lemma 15 (Main subfunction)**.** *For* $1.2 \leqslant x \leqslant 20$ *and* $-0.1 \leqslant y \leqslant 0.1$*,*

the function

$$e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{101}$$

is smaller than zero, is strictly monotonically increasing in x*, and strictly monotonically decreasing in* y *for the minimal* $x = 12/10 = 1.2$*.*

Proof. See proof 44. □

The graph of the subfunction in the specified domain is displayed in Figure A3.

Theorem 16 (Contraction ν-mapping)**.** *The mapping of the variance* $\tilde{\nu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$ *given in Eq. (5) is contracting for* $\lambda = \lambda_{01}$*,* $\alpha = \alpha_{01}$ *and the domain* Ω^+*:* $-0.1 \leqslant \mu \leqslant 0.1$*,* $-0.1 \leqslant \omega \leqslant 0.1$*,* $1.5 \leqslant \nu \leqslant 16$*, and* $0.8 \leqslant \tau \leqslant 1.25$*, that is,*

$$\left| \frac{\partial}{\partial \nu} \tilde{\nu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \right| \; < \; 1 \, . \tag{102}$$

Proof. In this domain Ω^+ we have the following three properties (see further below): $\frac{\partial}{\partial \nu} \tilde{\xi} < 1$, $\tilde{\mu} > 0$, and $\frac{\partial}{\partial \nu} \tilde{\mu} > 0$. Therefore, we have

$$\left| \frac{\partial}{\partial \nu} \tilde{\nu} \right| = \left| \frac{\partial}{\partial \nu} \tilde{\xi} - 2\tilde{\mu} \frac{\partial}{\partial \nu} \tilde{\mu} \right| < \left| \frac{\partial}{\partial \nu} \tilde{\xi} \right| < 1 \tag{103}$$

- We first proof that $\frac{\partial}{\partial \nu} \tilde{\xi} < 1$ in an even larger domain that fully contains Ω^+. According to Eq. (54), the derivative of the mapping Eq. (5) with respect to the variance ν is

$$\begin{aligned} &\frac{\partial}{\partial \nu} \tilde{\xi}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) \; = \\ &\frac{1}{2} \lambda^2 \tau \left(\alpha^2 \left(-e^{\mu\omega + \frac{\nu\tau}{2}} \right) \operatorname{erfc}\left(\frac{\mu\omega + \nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) + \right. \\ &\left. 2\alpha^2 e^{2\mu\omega + 2\nu\tau} \operatorname{erfc}\left(\frac{\mu\omega + 2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}} \right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}} \right) + 2 \right) \, . \end{aligned} \tag{104}$$

For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$, $-1 \leqslant \mu \leqslant 1$, $-0.1 \leqslant \omega \leqslant 0.1$ $1.5 \leqslant \nu \leqslant 16$, and $0.8 \leqslant \tau \leqslant 1.25$, we first show that the derivative is positive and then upper bound it.

According to Lemma 15, the expression

$$e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) \tag{105}$$

is negative. This expression multiplied by positive factors is subtracted in the derivative Eq. (104), therefore, the whole term is positive. The remaining term

$$2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) \tag{106}$$

of the derivative Eq. (104) is also positive according to Lemma 21. All factors outside the brackets in Eq. (104) are positive. Hence, the derivative Eq. (104) is positive.

The upper bound of the derivative is:

$$\begin{aligned}
&\frac{1}{2}\lambda_{01}^2\tau\left(\alpha_{01}^2\left(-e^{\mu\omega+\frac{\nu\tau}{2}}\right)\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)+\right. \\
&\left.2\alpha_{01}^2 e^{2\mu\omega+2\nu\tau}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\right) \;= \\
&\frac{1}{2}\lambda_{01}^2\tau\left(\alpha_{01}^2\left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)\left(e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-\right.\right. \\
&\left.\left.2e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\right) \;\leqslant \\
&\frac{1}{2}1.25\lambda_{01}^2\left(\alpha_{01}^2\left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)\left(e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-\right.\right. \\
&\left.\left.2e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\right) \;\leqslant \\
&\frac{1}{2}1.25\lambda_{01}^2\left(\alpha_{01}^2\left(e^{\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)-\right.\right. \\
&\left.\left.2e^{\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)\right)\left(-e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\right)-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\right) \;\leqslant \\
&\frac{1}{2}1.25\lambda_{01}^2\left(-e^{0.0}\alpha_{01}^2\left(e^{\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)-\right.\right. \\
&\left.\left.2e^{\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)\right)-\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)+2\right) \;\leqslant \\
&\frac{1}{2}1.25\lambda_{01}^2\left(-e^{0.0}\alpha_{01}^2\left(e^{\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)-\right.\right. \\
&\left.\left.2e^{\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)^2}\operatorname{erfc}\left(\frac{2\cdot 1.2+0.1}{\sqrt{2}\sqrt{1.2}}\right)\right)-\operatorname{erfc}\left(\frac{0.1}{\sqrt{2}\sqrt{1.2}}\right)+2\right) \;\leqslant \\
&0.995063 \;<\; 1\,.
\end{aligned} \tag{107}$$

We explain the chain of inequalities:

- First equality brings the expression into a shape where we can apply Lemma 15 for the the function Eq. (101).
- First inequality: The overall factor τ is bounded by 1.25.
- Second inequality: We apply Lemma 15. According to Lemma 15 the function Eq. (101) is negative. The largest contribution is to subtract the most negative value of the function Eq. (101), that is, the minimum of function Eq. (101). According to Lemma 15 the function Eq. (101) is strictly monotonically increasing in x and strictly monotonically decreasing in y for $x = 1.2$. Therefore the function Eq. (101) has its minimum at minimal $x = \nu\tau = 1.5 \cdot 0.8 = 1.2$ and maximal $y = \mu\omega = 1.0 \cdot 0.1 = 0.1$. We insert these values into the expression.

- Third inequality: We use for the whole expression the maximal factor $e^{-\frac{\mu^2\omega^2}{2\nu\tau}} < 1$ by setting this factor to 1.
- Fourth inequality: erfc is strictly monotonically decreasing. Therefore we maximize its argument to obtain the least value which is subtracted. We use the minimal $x = \nu\tau = 1.5 \cdot 0.8 = 1.2$ and the maximal $y = \mu\omega = 1.0 \cdot 0.1 = 0.1$.
- Sixth inequality: evaluation of the terms.

- We now show that $\tilde{\mu} > 0$. The expression $\tilde{\mu}(\mu, \omega, \nu, \tau)$ (Eq. (4)) is strictly monotonically increasing im $\mu\omega$ and $\nu\tau$. Therefore, the minimal value in Ω^+ is obtained at $\tilde{\mu}(0.01, 0.01, 1.5, 0.8) = 0.008293 > 0$.

- Last we show that $\frac{\partial}{\partial\nu}\tilde{\mu} > 0$. The expression $\frac{\partial}{\partial\nu}\tilde{\mu}(\mu, \omega, \nu, \tau) = \mathcal{J}_{12}(\mu, \omega, \nu, \tau)$ (Eq. (54)) can we reformulated as follows:

$$\mathcal{J}_{12}(\mu, \omega, \nu, \tau, \lambda, \alpha) = \frac{\lambda\tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\left(\sqrt{\pi}\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\mathrm{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \frac{\sqrt{2}(\alpha-1)}{\sqrt{\nu\tau}}\right)}{4\sqrt{\pi}} \tag{108}$$

is larger than zero when the term $\sqrt{\pi}\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\mathrm{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \frac{\sqrt{2}(\alpha-1)}{\sqrt{\nu\tau}}$ is larger than zero. This term obtains its minimal value at $\mu\omega = 0.01$ and $\nu\tau = 16 \cdot 1.25$, which can easily be shown using the Abramowitz bounds (Lemma 22) and evaluates to 0.16, therefore $\mathcal{J}_{12} > 0$ in Ω^+.

$\square$

A3.4.4 Lemmata for proofing Theorem 3: The variance is expanding

Main Sub-Function From Below. We consider functions in $\mu\omega$ and $\nu\tau$, therefore we set $x = \mu\omega$ and $y = \nu\tau$.

For $\lambda = \lambda_{01}$ and $\alpha = \alpha_{01}$, we consider the domain $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$ $0.00875 \leqslant \nu \leqslant 0.7$, and $0.8 \leqslant \tau \leqslant 1.25$.

For x and y we obtain: $0.8 \cdot 0.00875 = 0.007 \leqslant x \leqslant 0.875 = 1.25 \cdot 0.7$ and $0.1 \cdot (-0.1) = -0.01 \leqslant y \leqslant 0.01 = 0.1 \cdot 0.1$. In the following we assume to be within this domain.

In this domain, we consider the main sub-function of the derivate of second moment in the next layer, $J22$ (Eq. (54)):

$$\frac{\partial}{\partial\nu}\tilde{\xi} = \frac{1}{2}\lambda^2\tau\left(-\alpha^2 e^{\mu\omega+\frac{\nu\tau}{2}}\mathrm{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\alpha^2 e^{2\mu\omega+2\nu\tau}\mathrm{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right) \tag{109}$$

that depends on $\mu\omega$ and $\nu\tau$, therefore we set $x = \nu\tau$ and $y = \mu\omega$. Algebraic reformulations provide the formula in the following form:

$$\frac{\partial}{\partial\nu}\tilde{\xi} = \tag{110}$$
$$\frac{1}{2}\lambda^2\tau\left(\alpha^2\left(-e^{-\frac{y^2}{2x}}\right)\left(e^{\frac{(x+y)^2}{2x}}\mathrm{erfc}\left(\frac{y+x}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\mathrm{erfc}\left(\frac{y+2x}{\sqrt{2}\sqrt{x}}\right)\right) - \mathrm{erfc}\left(\frac{y}{\sqrt{2}\sqrt{x}}\right) + 2\right)$$

Lemma 17 (Main subfunction Below). *For $0.007 \leqslant x \leqslant 0.875$ and $-0.01 \leqslant y \leqslant 0.01$, the function*

$$e^{\frac{(x+y)^2}{2x}}\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{111}$$

smaller than zero, is strictly monotonically increasing in x and strictly monotonically increasing in y for the minimal $x = 0.007 = 0.00875 \cdot 0.8$, $x = 0.56 = 0.7 \cdot 0.8$, $x = 0.128 = 0.16 \cdot 0.8$, and $x = 0.216 = 0.24 \cdot 0.9$ (lower bound of 0.9 on τ).

Proof. See proof 45. $\square$

Lemma 18 (Monotone Derivative). *For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$, $0.00875 \leqslant \nu \leqslant 0.7$, and $0.8 \leqslant \tau \leqslant 1.25$. We are interested of the derivative of*

$$\tau \left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2e^{\left(\frac{\mu\omega+2\cdot\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) . \tag{112}$$

The derivative of the equation above with respect to

- *ν is larger than zero;*
- *τ is smaller than zero for maximal $\nu = 0.7$, $\nu = 0.16$, and $\nu = 0.24$ (with $0.9 \leqslant \tau$);*
- *$y = \mu\omega$ is larger than zero for $\nu\tau = 0.008750.8 = 0.007$, $\nu\tau = 0.70.8 = 0.56$, $\nu\tau = 0.160.8 = 0.128$, and $\nu\tau = 0.24 \cdot 0.9 = 0.216$.*

Proof. See proof 46. $\square$

A3.4.5 Computer-assisted proof details for main Lemma 12 in Section A3.4.1.

Error Analysis. We investigate the error propagation for the singular value (Eq. (61)) if the function arguments μ, ω, ν, τ suffer from numerical imprecisions up to ϵ. To this end, we first derive error propagation rules based on the mean value theorem and then we apply these rules to the formula for the singular value.

Lemma 19 (Mean value theorem). *For a real-valued function f which is differentiable in the closed interval $[a, b]$, there exists $t \in [0, 1]$ with*

$$f(\boldsymbol{a}) \ - \ f(\boldsymbol{b}) \ = \ \nabla f(\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a})) \ \cdot \ (\boldsymbol{a} \ - \ \boldsymbol{b}) \, . \tag{113}$$

It follows that for computation with error Δx, there exists a $t \in [0, 1]$ with

$$|f(\boldsymbol{x} + \Delta\boldsymbol{x}) \ - \ f(\boldsymbol{x})| \ \leqslant \ \|\nabla f(\boldsymbol{x} + t\Delta\boldsymbol{x})\| \ \|\Delta\boldsymbol{x}\| \ . \tag{114}$$

Therefore the increase of the norm of the error after applying function f is bounded by the norm of the gradient $\|\nabla f(\boldsymbol{x} + t\Delta\boldsymbol{x})\|$.

We now compute for the functions, that we consider their gradient and its 2-norm:

- addition:
  $f(\boldsymbol{x}) = x_1 + x_2$ and $\nabla f(\boldsymbol{x}) = (1, 1)$, which gives $\|\nabla f(\boldsymbol{x})\| = \sqrt{2}$.
  We further know that
  $$|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x})| \ = \ |x_1 + x_2 + \Delta x_1 + \Delta x_2 - x_1 - x_2| \ \leqslant \ |\Delta x_1| + |\Delta x_2| \ . \tag{115}$$
  Adding n terms gives:
  $$\left|\sum_{i=1}^{n} x_i + \Delta x_i \ - \ \sum_{i=1}^{n} x_i\right| \ \leqslant \ \sum_{i=1}^{n} |\Delta x_i| \ \leqslant \ n \, |\Delta x_i|_{\max} \ . \tag{116}$$
- subtraction:
  $f(\boldsymbol{x}) = x_1 - x_2$ and $\nabla f(\boldsymbol{x}) = (1, -1)$, which gives $\|\nabla f(\boldsymbol{x})\| = \sqrt{2}$.
  We further know that
  $$|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x})| \ = \ |x_1 - x_2 + \Delta x_1 - \Delta x_2 - x_1 + x_2| \ \leqslant \ |\Delta x_1| + |\Delta x_2| \ . \tag{117}$$
  Subtracting n terms gives:
  $$\left|\sum_{i=1}^{n} -(x_i + \Delta x_i) \ + \ \sum_{i=1}^{n} x_i\right| \ \leqslant \ \sum_{i=1}^{n} |\Delta x_i| \ \leqslant \ n \, |\Delta x_i|_{\max} \ . \tag{118}$$

- multiplication:
  $f(\boldsymbol{x}) = x_1 x_2$ and $\nabla f(\boldsymbol{x}) = (x_2, x_1)$, which gives $\|\nabla f(\boldsymbol{x})\| = \|\boldsymbol{x}\|$.
  We further know that

$$|f(\boldsymbol{x}+\Delta\boldsymbol{x}) - f(\boldsymbol{x})| \;=\; |x_1 \cdot x_2 + \Delta x_1 \cdot x_2 + \Delta x_2 \cdot x_1 + \Delta x_1 \cdot \Delta x_s - x_1 \cdot x_2| \;\leqslant \tag{119}$$
$$|\Delta x_1|\,|x_2| + |\Delta x_2|\,|x_1| + O(\Delta^2)\ .$$

  Multiplying n terms gives:

$$\left|\prod_{i=1}^{n}(x_i + \Delta x_i) \;-\; \prod_{i=1}^{n} x_i\right| \;=\; \left|\prod_{i=1}^{n} x_i \sum_{i=1}^{n} \frac{\Delta x_i}{x_i} \;+\; O(\Delta^2)\right| \;\leqslant \tag{120}$$
$$\prod_{i=1}^{n} |x_i| \sum_{i=1}^{n} \left|\frac{\Delta x_i}{x_i}\right| \;+\; O(\Delta^2) \;\leqslant\; n \prod_{i=1}^{n} |x_i| \;\left|\frac{\Delta x_i}{x_i}\right|_{\max} \;+\; O(\Delta^2)\ .$$

- division:
  $f(\boldsymbol{x}) = \frac{x_1}{x_2}$ and $\nabla f(\boldsymbol{x}) = \left(\frac{1}{x_2}, -\frac{x_1}{x_2^2}\right)$, which gives $\|\nabla f(\boldsymbol{x})\| = \frac{\|\boldsymbol{x}\|}{x_2^2}$.
  We further know that

$$|f(\boldsymbol{x}+\Delta\boldsymbol{x}) - f(\boldsymbol{x})| \;=\; \left|\frac{x_1 + \Delta x_1}{x_2 + \Delta x_2} - \frac{x_1}{x_2}\right| \;=\; \left|\frac{(x_1 + \Delta x_1)x_2 - x_1(x_2 + \Delta x_2)}{(x_2 + \Delta x_2)x_2}\right| \;= \tag{121}$$
$$\left|\frac{\Delta x_1 \cdot x_2 - \Delta x_2 \cdot x_1}{x_2^2 + \Delta x_2 \cdot x_2}\right| \;=\; \left|\frac{\Delta x_1}{x_2} - \frac{\Delta x_2 \cdot x_1}{x_2^2}\right| + O(\Delta^2)\ .$$

- square root:
  $f(x) = \sqrt{x}$ and $f'(x) = \frac{1}{2\sqrt{x}}$, which gives $|f'(x)| = \frac{1}{2\sqrt{x}}$.
- exponential function:
  $f(x) = \exp(x)$ and $f'(x) = \exp(x)$, which gives $|f'(x)| = \exp(x)$.
- error function:
  $f(x) = \mathrm{erf}(x)$ and $f'(x) = \frac{2}{\sqrt{\pi}}\exp(-x^2)$, which gives $|f'(x)| = \frac{2}{\sqrt{\pi}}\exp(-x^2)$.
- complementary error function:
  $f(x) = \mathrm{erfc}(x)$ and $f'(x) = -\frac{2}{\sqrt{\pi}}\exp(-x^2)$, which gives $|f'(x)| = \frac{2}{\sqrt{\pi}}\exp(-x^2)$.

Lemma 20. *If the values μ, ω, ν, τ have a precision of ϵ, the singular value (Eq. (61)) evaluated with the formulas given in Eq. (54) and Eq. (61) has a precision better than 292ϵ.*

This means for a machine with a typical precision of $2^{-52} = 2.220446 \cdot 10^{-16}$, we have the rounding error $\epsilon \approx 10^{-16}$, the evaluation of the singular value (Eq. (61)) with the formulas given in Eq. (54) and Eq. (61) has a precision better than $10^{-13} > 292\epsilon$.

Proof. We have the numerical precision ϵ of the parameters μ, ω, ν, τ, that we denote by $\Delta\mu, \Delta\omega, \Delta\nu, \Delta\tau$ together with our domain Ω.

With the error propagation rules that we derived in Subsection A3.4.5, we can obtain bounds for the numerical errors on the following simple expressions:

$$\begin{aligned}
\Delta(\mu\omega) &\leqslant \Delta\mu\,|\omega| + \Delta\omega\,|\mu| \leqslant 0.2\epsilon \\
\Delta(\nu\tau) &\leqslant \Delta\nu\,|\tau| + \Delta\tau\,|\nu| \leqslant 1.5\epsilon + 1.5\epsilon = 3\epsilon \\
\Delta\left(\frac{\nu\tau}{2}\right) &\leqslant (\Delta(\nu\tau)2 + \Delta 2\,|\nu\tau|)\,\frac{1}{2^2} \leqslant (6\epsilon + 1.25 \cdot 1.5\epsilon)/4 < 2\epsilon \\
\Delta(\mu\omega + \nu\tau) &\leqslant \Delta(\mu\omega) + \Delta(\nu\tau) = 3.2\epsilon \\
\Delta\left(\mu\omega + \frac{\nu\tau}{2}\right) &\leqslant \Delta(\mu\omega) + \Delta\left(\frac{\nu\tau}{2}\right) < 2.2\epsilon \\
\Delta\left(\sqrt{\nu\tau}\right) &\leqslant \frac{\Delta(\nu\tau)}{2\sqrt{\nu\tau}} \leqslant \frac{3\epsilon}{2\sqrt{0.64}} = 1.875\epsilon
\end{aligned} \tag{122}$$

$$
\begin{aligned}
\Delta\left(\sqrt{2}\right) &\leqslant \frac{\Delta 2}{2\sqrt{2}} \leqslant \frac{1}{2\sqrt{2}}\epsilon \\
\Delta\left(\sqrt{2}\sqrt{\nu\tau}\right) &\leqslant \sqrt{2}\Delta\left(\sqrt{\nu\tau}\right) + \nu\tau\Delta\left(\sqrt{2}\right) \leqslant \sqrt{2}\cdot 1.875\epsilon + 1.5\cdot 1.25\cdot \frac{1}{2\sqrt{2}}\epsilon < 3.5\epsilon \\
\Delta\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) &\leqslant \left(\Delta\left(\mu\omega\right)\sqrt{2}\sqrt{\nu\tau} + |\mu\omega|\,\Delta\left(\sqrt{2}\sqrt{\nu\tau}\right)\right)\frac{1}{\left(\sqrt{2}\sqrt{\nu\tau}\right)^2} \leqslant \\
&\left(0.2\epsilon\sqrt{2}\sqrt{0.64} + 0.01\cdot 3.5\epsilon\right)\frac{1}{2\cdot 0.64} < 0.25\epsilon \\
\Delta\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) &\leqslant \left(\Delta\left(\mu\omega+\nu\tau\right)\sqrt{2}\sqrt{\nu\tau} + |\mu\omega+\nu\tau|\,\Delta\left(\sqrt{2}\sqrt{\nu\tau}\right)\right)\frac{1}{\left(\sqrt{2}\sqrt{\nu\tau}\right)^2} \leqslant \\
&\left(3.2\epsilon\sqrt{2}\sqrt{0.64} + 1.885\cdot 3.5\epsilon\right)\frac{1}{2\cdot 0.64} < 8\epsilon.
\end{aligned}
$$

Using these bounds on the simple expressions, we can now calculate bounds on the numerical errors of compound expressions:

$$
\begin{aligned}
\Delta\left(\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) &\leqslant \frac{2}{\sqrt{\pi}}e^{-\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\Delta\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) < \\
&\frac{2}{\sqrt{\pi}}0.25\epsilon < 0.3\epsilon
\end{aligned} \tag{123}
$$

$$
\begin{aligned}
\Delta\left(\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) &\leqslant \frac{2}{\sqrt{\pi}}e^{-\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\Delta\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) < \\
&\frac{2}{\sqrt{\pi}}8\epsilon < 10\epsilon
\end{aligned} \tag{124}
$$

$$
\Delta\left(e^{\mu\omega+\frac{\nu\tau}{2}}\right) \leqslant \left(e^{\mu\omega+\frac{\nu\tau}{2}}\right)\Delta\left(e^{\mu\omega+\frac{\nu\tau}{2}}\right) < \tag{125}
$$

$$
e^{0.9475}2.2\epsilon < 5.7\epsilon \tag{126}
$$

Subsequently, we can use the above results to get bounds for the numerical errors on the Jacobian entries (Eq. (54)), applying the rules from Subsection A3.4.5 again:

$$
\Delta\left(\mathcal{J}_{11}\right) = \Delta\left(\frac{1}{2}\lambda\omega\left(\alpha e^{\mu\omega+\frac{\nu\tau}{2}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + 2\right)\right) < 6\epsilon, \tag{127}
$$

and we obtain $\Delta\left(\mathcal{J}_{12}\right) < 78\epsilon$, $\Delta\left(\mathcal{J}_{21}\right) < 189\epsilon$, $\Delta\left(\mathcal{J}_{22}\right) < 405\epsilon$ and $\Delta\left(\tilde{\mu}\right) < 52\epsilon$. We also have bounds on the absolute values on $\mathcal{J}_{ij}$ and $\tilde{\mu}$ (see Lemma 6, Lemma 7, and Lemma 9), therefore we can propagate the error also through the function that calculates the singular value (Eq. (61)).

$$
\begin{aligned}
&\Delta\left(S(\mu,\omega,\nu,\tau,\lambda,\alpha)\right) = \\
&\Delta\left(\frac{1}{2}\left(\sqrt{(\mathcal{J}_{11}+\mathcal{J}_{22}-2\tilde{\mu}\mathcal{J}_{12})^2+(\mathcal{J}_{21}-2\tilde{\mu}\mathcal{J}_{11}-\mathcal{J}_{12})^2} + \right.\right. \\
&\left.\left.\sqrt{(\mathcal{J}_{11}-\mathcal{J}_{22}+2\tilde{\mu}\mathcal{J}_{12})^2+(\mathcal{J}_{12}+\mathcal{J}_{21}-2\tilde{\mu}\mathcal{J}_{11})^2}\right)\right) < 292\epsilon.
\end{aligned} \tag{128}
$$

□

Precision of Implementations. We will show that our computations are correct up to 3 ulps. For our implementation in GNU C library and the hardware architectures that we used, the precision of all mathematical functions that we used is at least one ulp. The term "ulp" (acronym for "unit in the last place") was coined by W. Kahan in 1960. It is the highest precision (up to some factor smaller 1), which can be achieved for the given hardware and floating point representation.

Kahan defined ulp as [21]:

> "Ulp(x) is the gap between the two *finite* floating-point numbers nearest x, even if x is one of them. (But ulp(NaN) is NaN.)"

Harrison defined ulp as [15]:

> "an ulp in x is the distance between the two closest *straddling* floating point numbers a and b, i.e. those with $a \leqslant x \leqslant b$ and $a \neq b$ assuming an unbounded exponent range."

In the literature we find also slightly different definitions [29].

According to [29] who refers to [11]:

> "IEEE-754 mandates four standard rounding modes:"
>
> "Round-to-nearest: $r(x)$ is the floating-point value closest to x with the usual distance; if two floating-point value are equally close to x, then $r(x)$ is the one whose least significant bit is equal to zero."
>
> "IEEE-754 standardises 5 operations: addition (which we shall note $\oplus$ in order to distinguish it from the operation over the reals), subtraction ($\ominus$), multiplication ($\otimes$), division ($\oslash$), and also square root."
>
> "IEEE-754 specifies em exact rounding [Goldberg, 1991, §1.5]: the result of a floating-point operation is the same as if the operation were performed on the real numbers with the given inputs, then rounded according to the rules in the preceding section. Thus, $x \oplus y$ is defined as $r(x + y)$, with x and y taken as elements of $\mathbb{R} \cup \{-\infty, +\infty\}$; the same applies for the other operators."

Consequently, the IEEE-754 standard guarantees that addition, subtraction, multiplication, division, and squared root is precise up to one ulp.

We have to consider transcendental functions. First the is the exponential function, and then the complementary error function $\mathrm{erfc}(x)$, which can be computed via the error function $\mathrm{erf}(x)$.

Intel states [29]:

> "With the Intel486 processor and Intel 387 math coprocessor, the worst- case, transcendental function error is typically 3 or 3.5 ulps, but is some- times as large as 4.5 ulps."

According to `https://www.mirbsd.org/htman/i386/man3/exp.htm` and `http://man.openbsd.org/OpenBSD-current/man3/exp.3`:

> "$\exp(x)$, $\log(x)$, $\mathrm{expm1}(x)$ and $\mathrm{log1p}(x)$ are accurate to within an ulp"

which is the same for freebsd `https://www.freebsd.org/cgi/man.cgi?query=exp&sektion=3&apropos=0&manpath=freebsd`:

> "The values of exp(0), expm1(0), exp2(integer), and pow(integer, integer) are exact provided that they are representable. Otherwise the error in these functions is generally below one ulp."

The same holds for "FDLIBM" `http://www.netlib.org/fdlibm/readme`:

> "FDLIBM is intended to provide a reasonably portable (see assumptions below), reference quality (below one ulp for major functions like sin,cos,exp,log) math library (libm.a)."

In `http://www.gnu.org/software/libc/manual/html_node/Errors-in-Math-Functions.html` we find that both $\exp$ and erf have an error of 1 ulp while erfc has an error up to 3 ulps depending on the architecture. For the most common architectures as used by us, however, the error of erfc is 1 ulp.

We implemented the function in the programming language C. We rely on the GNU C Library [26]. According to the GNU C Library manual which can be obtained from `http://www.gnu.org/`

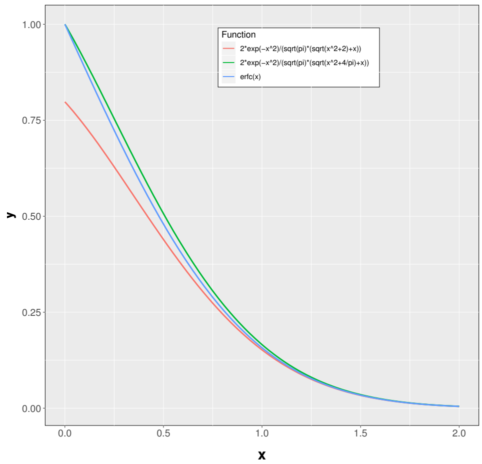

Figure A4: Graphs of the upper and lower bounds on erfc. The lower bound $\frac{2e^{-x^2}}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)}$ (red), the upper bound $\frac{2e^{-x^2}}{\sqrt{\pi}\left(\sqrt{x^2+\frac{4}{\pi}}+x\right)}$ (green) and the function $\mathrm{erfc}(x)$ (blue) as treated in Lemma 22.

`software/libc/manual/pdf/libc.pdf`, the errors of the math functions $\exp$, erf, and erfc are not larger than 3 ulps for all architectures [26, pp. 528]. For the architectures ix86, i386/i686/fpu, and m68k/fpmu68k/m680x0/fpu that we used the error are at least one ulp [26, pp. 528].

A3.4.6 Intermediate Lemmata and Proofs

Since we focus on the fixed point $(\mu, \nu) = (0, 1)$, we assume for our whole analysis that $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. Furthermore, we restrict the range of the variables $\mu \in [\mu_{\min}, \mu_{\max}] = [-0.1, 0.1]$, $\omega \in [\omega_{\min}, \omega_{\max}] = [-0.1, 0.1]$, $\nu \in [\nu_{\min}, \nu_{\max}] = [0.8, 1.5]$, and $\tau \in [\tau_{\min}, \tau_{\max}] = [0.8, 1.25]$.

For bounding different partial derivatives we need properties of different functions. We will bound a the absolute value of a function by computing an upper bound on its maximum and a lower bound on its minimum. These bounds are computed by upper or lower bounding terms. The bounds get tighter if we can combine terms to a more complex function and bound this function. The following lemmata give some properties of functions that we will use in bounding complex functions.

Throughout this work, we use the error function $\mathrm{erf}(x) := \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2}$ and the complementary error function $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$.

Lemma 21 (Basic functions)**.** *$\exp(x)$ is strictly monotonically increasing from 0 at $-\infty$ to ∞ at ∞ and has positive curvature.*

According to its definition $\mathrm{erfc}(x)$ is strictly monotonically decreasing from 2 at $-\infty$ to 0 at ∞.

Next we introduce a bound on erfc:

Lemma 22 (Erfc bound from Abramowitz)**.**

$$\frac{2e^{-x^2}}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} < \mathrm{erfc}(x) \leqslant \frac{2e^{-x^2}}{\sqrt{\pi}\left(\sqrt{x^2+\frac{4}{\pi}}+x\right)}, \tag{129}$$

for $x > 0$.

Proof. The statement follows immediately from [1] (page 298, formula 7.1.13). □

These bounds are displayed in figure A4.

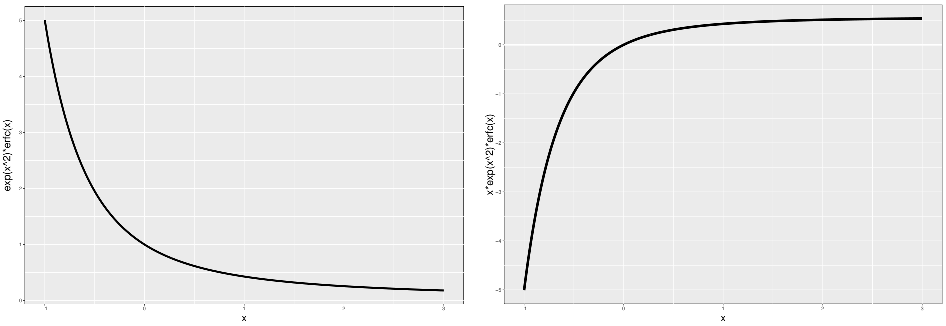

Figure A5: Graphs of the functions $e^{x^2}\,\mathrm{erfc}(x)$ (left) and $xe^{x^2}\,\mathrm{erfc}(x)$ (right) treated in Lemma 23 and Lemma 24, respectively.

Lemma 23 (Function $e^{x^2}\,\mathrm{erfc}(x)$)**.** *$e^{x^2}\,\mathrm{erfc}(x)$ is strictly monotonically decreasing for $x > 0$ and has positive curvature (positive 2nd order derivative), that is, the decreasing slowes down.*

A graph of the function is displayed in Figure A5.

Proof. The derivative of $e^{x^2}\,\mathrm{erfc}(x)$ is

$$\frac{\partial e^{x^2}\,\mathrm{erfc}(x)}{\partial x} = 2e^{x^2} x\,\mathrm{erfc}(x) - \frac{2}{\sqrt{\pi}} \,. \tag{130}$$

Using Lemma 22, we get

$$\frac{\partial e^{x^2}\,\mathrm{erfc}(x)}{\partial x} = 2e^{x^2} x\,\mathrm{erfc}(x) - \frac{2}{\sqrt{\pi}} < \frac{4x}{\sqrt{\pi}\left(\sqrt{x^2+\frac{4}{\pi}}+x\right)} - \frac{2}{\sqrt{\pi}} = \frac{2\left(\frac{2}{\sqrt{\frac{4}{\pi x^2}+1}+1} - 1\right)}{\sqrt{\pi}} < 0 \tag{131}$$

Thus $e^{x^2}\,\mathrm{erfc}(x)$ is strictly monotonically decreasing for $x > 0$.

The second order derivative of $e^{x^2}\,\mathrm{erfc}(x)$ is

$$\frac{\partial^2 e^{x^2}\,\mathrm{erfc}(x)}{\partial x^2} = 4e^{x^2} x^2\,\mathrm{erfc}(x) + 2e^{x^2}\,\mathrm{erfc}(x) - \frac{4x}{\sqrt{\pi}} \,. \tag{132}$$

Again using Lemma 22 (first inequality), we get

$$\begin{aligned} & 2\left(\left(2x^2+1\right) e^{x^2}\,\mathrm{erfc}(x) - \frac{2x}{\sqrt{\pi}}\right) > \\ & \frac{4\left(2x^2+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} - \frac{4x}{\sqrt{\pi}} = \\ & \frac{4\left(x^2-\sqrt{x^2+2}x+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} = \\ & \frac{4\left(x^2-\sqrt{x^4+2x^2}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} > \\ & \frac{4\left(x^2-\sqrt{x^4+2x^2+1}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} = 0 \end{aligned} \tag{133}$$

For the last inequality we added 1 in the numerator in the square root which is subtracted, that is, making a larger negative term in the numerator. ∎

Lemma 24 (Properties of $xe^{x^2}\operatorname{erfc}(x)$)**.** *The function $xe^{x^2}\operatorname{erfc}(x)$ has the sign of x and is monotonically increasing to $\frac{1}{\sqrt{\pi}}$.*

Proof. The derivative of $xe^{x^2}\operatorname{erfc}(x)$ is

$$2e^{x^2}x^2\operatorname{erfc}(x) + e^{x^2}\operatorname{erfc}(x) - \frac{2x}{\sqrt{\pi}} \,. \tag{134}$$

This derivative is positive since

$$\begin{aligned}
&2e^{x^2}x^2\operatorname{erfc}(x) + e^{x^2}\operatorname{erfc}(x) - \frac{2x}{\sqrt{\pi}} \ = \\
&e^{x^2}\left(2x^2+1\right)\operatorname{erfc}(x) - \frac{2x}{\sqrt{\pi}} > \frac{2\left(2x^2+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} - \frac{2x}{\sqrt{\pi}} = \frac{2\left(\left(2x^2+1\right)-x\left(\sqrt{x^2+2}+x\right)\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} \ = \\
&\frac{2\left(x^2-x\sqrt{x^2+2}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} = \frac{2\left(x^2-x\sqrt{x^2+2}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} > \frac{2\left(x^2-x\sqrt{x^2+\frac{1}{x^2}+2}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} \ = \\
&\frac{2\left(x^2-\sqrt{x^4+2x^2+1}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} = \frac{2\left(x^2-\sqrt{\left(x^2+1\right)^2}+1\right)}{\sqrt{\pi}\left(\sqrt{x^2+2}+x\right)} = 0 \,.
\end{aligned} \tag{135}$$

We apply Lemma 22 to $x\operatorname{erfc}(x)e^{x^2}$ and divide the terms of the lemma by x, which gives

$$\frac{2}{\sqrt{\pi}\left(\sqrt{\frac{2}{x^2}+1}+1\right)} < x\operatorname{erfc}(x)e^{x^2} \leqslant \frac{2}{\sqrt{\pi}\left(\sqrt{\frac{4}{\pi x^2}+1}+1\right)} \,. \tag{136}$$

For $\lim_{x\to\infty}$ both the upper and the lower bound go to $\frac{1}{\sqrt{\pi}}$. ☐

Lemma 25 (Function $\mu\omega$)**.** *$h_{11}(\mu,\omega) = \mu\omega$ is monotonically increasing in $\mu\omega$. It has minimal value $t_{11} = -0.01$ and maximal value $T_{11} = 0.01$.*

Proof. Obvious. ☐

Lemma 26 (Function $\nu\tau$)**.** *$h_{22}(\nu,\tau) = \nu\tau$ is monotonically increasing in $\nu\tau$ and is positive. It has minimal value $t_{22} = 0.64$ and maximal value $T_{22} = 1.875$.*

Proof. Obvious. ☐

Lemma 27 (Function $\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}$)**.** *$h_1(\mu,\omega,\nu,\tau) = \frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}$ is larger than zero and increasing in both $\nu\tau$ and $\mu\omega$. It has minimal value $t_1 = 0.5568$ and maximal value $T_1 = 0.9734$.*

Proof. The derivative of the function $\frac{\mu\omega+x}{\sqrt{2}\sqrt{x}}$ with respect to x is

$$\frac{1}{\sqrt{2}\sqrt{x}} - \frac{\mu\omega+x}{2\sqrt{2}x^{3/2}} = \frac{2x-(\mu\omega+x)}{2\sqrt{2}x^{3/2}} = \frac{x-\mu\omega}{2\sqrt{2}x^{3/2}} \ > \ 0 \,, \tag{137}$$

since $x > 0.8\cdot 0.8$ and $\mu\omega < 0.1\cdot 0.1$. ☐

Lemma 28 (Function $\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}$)**.** *$h_2(\mu,\omega,\nu,\tau) = \frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}$ is larger than zero and increasing in both $\nu\tau$ and $\mu\omega$. It has minimal value $t_2 = 1.1225$ and maximal value $T_2 = 1.9417$.*

Proof. The derivative of the function $\frac{\mu\omega+2x}{\sqrt{2}\sqrt{x}}$ with respect to x is

$$\frac{\sqrt{2}}{\sqrt{x}} - \frac{\mu\omega+2x}{2\sqrt{2}x^{3/2}} = \frac{4x-(\mu\omega+2x)}{2\sqrt{2}x^{3/2}} = \frac{2x-\mu\omega}{2\sqrt{2}x^{3/2}} \ > \ 0 \,. \tag{138}$$

☐

Lemma 29 (Function $\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}$). *$h_3(\mu,\omega,\nu,\tau) = \frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}$ monotonically decreasing in $\nu\tau$ and monotonically increasing in $\mu\omega$. It has minimal value $t_3 = -0.0088388$ and maximal value $T_3 = 0.0088388$.*

Proof. Obvious. $\square$

Lemma 30 (Function $\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)^2$). *$h_4(\mu,\omega,\nu,\tau) = \left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)^2$ has a minimum at 0 for $\mu = 0$ or $\omega = 0$ and has a maximum for the smallest $\nu\tau$ and largest $|\mu\omega|$ and is larger or equal to zero. It has minimal value $t_4 = 0$ and maximal value $T_4 = 0.000078126$.*

Proof. Obvious. $\square$

Lemma 31 (Function $\frac{\sqrt{\frac{2}{\pi}}(\alpha-1)}{\sqrt{\nu\tau}}$). *$\frac{\sqrt{\frac{2}{\pi}}(\alpha-1)}{\sqrt{\nu\tau}} > 0$ and decreasing in $\nu\tau$.*

Proof. Statements follow directly from elementary functions square root and division. $\square$

Lemma 32 (Function $2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)$). *$2 - \operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) > 0$ and decreasing in $\nu\tau$ and increasing in $\mu\omega$.*

Proof. Statements follow directly from Lemma 21 and erfc. $\square$

Lemma 33 (Function $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha-1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}}\right)$). *For $\lambda = \lambda_{01}$ and $\alpha = \alpha_{01}$, $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha-1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}}\right) < 0$ and increasing in both $\nu\tau$ and $\mu\omega$.*

Proof. We consider the function $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha-1)\mu\omega}{x^{3/2}} - \frac{\alpha}{\sqrt{x}}\right)$, which has the derivative with respect to x:

$$\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{2x^{3/2}} - \frac{3(\alpha-1)\mu\omega}{2x^{5/2}}\right) . \tag{139}$$

This derivative is larger than zero, since

$$\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{2(\nu\tau)^{3/2}} - \frac{3(\alpha-1)\mu\omega}{2(\nu\tau)^{5/2}}\right) > \frac{\sqrt{\frac{2}{\pi}}\left(\alpha - \frac{3(\alpha-1)\mu\omega}{\nu\tau}\right)}{2(\nu\tau)^{3/2}} > 0 . \tag{140}$$

The last inequality follows from $\alpha - \frac{3 \cdot 0.1 \cdot 0.1(\alpha-1)}{0.8 \cdot 0.8} > 0$ for $\alpha = \alpha_{01}$.

We next consider the function $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha-1)x}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}}\right)$, which has the derivative with respect to x:

$$\frac{\sqrt{\frac{2}{\pi}}(\alpha-1)}{(\nu\tau)^{3/2}} > 0 . \tag{141}$$

$\square$

Lemma 34 (Function $\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha-1)\mu^2\omega^2}{(\nu\tau)^{3/2}} + \frac{-\alpha+\alpha\mu\omega+1}{\sqrt{\nu\tau}} - \alpha\sqrt{\nu\tau}\right)$). *The function $\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha-1)\mu^2\omega^2}{(\nu\tau)^{3/2}} + \frac{-\alpha+\alpha\mu\omega+1}{\sqrt{\nu\tau}} - \alpha\sqrt{\nu\tau}\right) < 0$ is decreasing in $\nu\tau$ and increasing in $\mu\omega$.*

Proof. We define the function

$$\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha-1)\mu^2\omega^2}{x^{3/2}} + \frac{-\alpha+\alpha\mu\omega+1}{\sqrt{x}} - \alpha\sqrt{x}\right) \tag{142}$$

which has as derivative with respect to x:

$$\sqrt{\frac{2}{\pi}}\left(\frac{3(\alpha-1)\mu^2\omega^2}{2x^{5/2}} - \frac{-\alpha+\alpha\mu\omega+1}{2x^{3/2}} - \frac{\alpha}{2\sqrt{x}}\right) = \tag{143}$$

$$\frac{1}{\sqrt{2\pi}x^{5/2}}\left(3(\alpha-1)\mu^2\omega^2-x(-\alpha+\alpha\mu\omega+1)-\alpha x^2\right) \ .$$

The derivative of the term $3(\alpha-1)\mu^2\omega^2-x(-\alpha+\alpha\mu\omega+1)-\alpha x^2$ with respect to x is $-1+\alpha-\mu\omega\alpha-2\alpha x<0$, since $2\alpha x>1.6\alpha$. Therefore the term is maximized with the smallest value for x, which is $x=\nu\tau=0.8\cdot 0.8$. For $\mu\omega$ we use for each term the value which gives maximal contribution. We obtain an upper bound for the term:

$$3(-0.1\cdot 0.1)^2(\alpha_{01}-1)-(0.8\cdot 0.8)^2\alpha_{01}-0.8\cdot 0.8((-0.1\cdot 0.1)\alpha_{01}-\alpha_{01}+1) \ = \ -0.243569 \ . \tag{144}$$

Therefore the derivative with respect to $x=\nu\tau$ is smaller than zero and the original function is decreasing in $\nu\tau$

We now consider the derivative with respect to $x=\mu\omega$. The derivative with respect to x of the function

$$\sqrt{\frac{2}{\pi}}\left(-\alpha\sqrt{\nu\tau}-\frac{(\alpha-1)x^2}{(\nu\tau)^{3/2}}+\frac{-\alpha+\alpha x+1}{\sqrt{\nu\tau}}\right) \tag{145}$$

is

$$\frac{\sqrt{\frac{2}{\pi}}(\alpha\nu\tau-2(\alpha-1)x)}{(\nu\tau)^{3/2}} \ . \tag{146}$$

Since $-2x(-1+\alpha)+\nu\tau\alpha>-2\cdot 0.01\cdot(-1+\alpha_{01})+0.8\cdot 0.8\alpha_{01}>1.0574>0$, the derivative is larger than zero. Consequently, the original function is increasing in $\mu\omega$.

The maximal value is obtained with the minimal $\nu\tau=0.8\cdot 0.8$ and the maximal $\mu\omega=0.1\cdot 0.1$. The maximal value is

$$\sqrt{\frac{2}{\pi}}\left(\frac{0.1\cdot 0.1\alpha_{01}-\alpha_{01}+1}{\sqrt{0.8\cdot 0.8}}+\frac{0.1^2 0.1^2(-1)(\alpha_{01}-1)}{(0.8\cdot 0.8)^{3/2}}-\sqrt{0.8\cdot 0.8}\alpha_{01}\right) \ = \ -1.72296 \ . \tag{147}$$

Therefore the original function is smaller than zero. ☐

Lemma 35 (Function $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{(\nu\tau)^{3/2}}-\frac{3\alpha^2}{\sqrt{\nu\tau}}\right)$)**.** *For $\lambda=\lambda_{01}$ and $\alpha=\alpha_{01}$, $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{(\nu\tau)^{3/2}}-\frac{3\alpha^2}{\sqrt{\nu\tau}}\right)<0$ and increasing in both $\nu\tau$ and $\mu\omega$.*

Proof. The derivative of the function

$$\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha^2-1\right)\mu\omega}{x^{3/2}}-\frac{3\alpha^2}{\sqrt{x}}\right) \tag{148}$$

with respect to x is

$$\sqrt{\frac{2}{\pi}}\left(\frac{3\alpha^2}{2x^{3/2}}-\frac{3\left(\alpha^2-1\right)\mu\omega}{2x^{5/2}}\right) \ = \ \frac{3\left(\alpha^2x-\left(\alpha^2-1\right)\mu\omega\right)}{\sqrt{2\pi}x^{5/2}} \ > \ 0 \ , \tag{149}$$

since $\alpha^2x-\mu\omega(-1+\alpha^2)>\alpha_{01}^2 0.8\cdot 0.8-0.1\cdot 0.1\cdot(-1+\alpha_{01}^2)>1.77387$

The derivative of the function

$$\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha^2-1\right)x}{(\nu\tau)^{3/2}}-\frac{3\alpha^2}{\sqrt{\nu\tau}}\right) \tag{150}$$

with respect to x is

$$\frac{\sqrt{\frac{2}{\pi}}\left(\alpha^2-1\right)}{(\nu\tau)^{3/2}} \ > \ 0 \ . \tag{151}$$

The maximal function value is obtained by maximal $\nu\tau=1.5\cdot 1.25$ and the maximal $\mu\omega=0.1\cdot 0.1$. The maximal value is $\sqrt{\frac{2}{\pi}}\left(\frac{0.1\cdot 0.1\left(\alpha_{01}^2-1\right)}{(1.5\cdot 1.25)^{3/2}}-\frac{3\alpha_{01}^2}{\sqrt{1.5\cdot 1.25}}\right) \ = \ -4.88869$. Therefore the function is negative. ☐

Lemma 36 (Function $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{\sqrt{\nu\tau}} - 3\alpha^2\sqrt{\nu\tau}\right)$**).** *The function* $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{\sqrt{\nu\tau}} - 3\alpha^2\sqrt{\nu\tau}\right) < 0$ *is decreasing in* $\nu\tau$ *and increasing in* $\mu\omega$*.*

Proof. The derivative of the function

$$\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha^2-1\right)\mu\omega}{\sqrt{x}} - 3\alpha^2\sqrt{x}\right) \tag{152}$$

with respect to x is

$$\sqrt{\frac{2}{\pi}}\left(-\frac{\left(\alpha^2-1\right)\mu\omega}{2x^{3/2}} - \frac{3\alpha^2}{2\sqrt{x}}\right) \;=\; \frac{-\left(\alpha^2-1\right)\mu\omega - 3\alpha^2 x}{\sqrt{2\pi}x^{3/2}} \;<\; 0\,, \tag{153}$$

since $-3\alpha^2 x - \mu\omega(-1+\alpha^2) < -3\alpha_{01}^2 0.8\cdot 0.8 + 0.1\cdot 0.1(-1+\alpha_{01}^2) < -5.35764$.
The derivative of the function

$$\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha^2-1\right)x}{\sqrt{\nu\tau}} - 3\alpha^2\sqrt{\nu\tau}\right) \tag{154}$$

with respect to x is

$$\frac{\sqrt{\frac{2}{\pi}}\left(\alpha^2-1\right)}{\sqrt{\nu\tau}} \;>\; 0\,. \tag{155}$$

The maximal function value is obtained for minimal $\nu\tau = 0.8\cdot 0.8$ and the maximal $\mu\omega = 0.1\cdot 0.1$. The value is $\sqrt{\frac{2}{\pi}}\left(\frac{0.1\cdot 0.1\left(\alpha_{01}^2-1\right)}{\sqrt{0.8\cdot 0.8}} - 3\sqrt{0.8\cdot 0.8}\alpha_{01}^2\right) = -5.34347$. Thus, the function is negative. □

Lemma 37 (Function $\nu\tau e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$**).** *The function* $\nu\tau e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) > 0$ *is increasing in* $\nu\tau$ *and decreasing in* $\mu\omega$*.*

Proof. The derivative of the function

$$x e^{\frac{(\mu\omega+x)^2}{2x}}\operatorname{erfc}\left(\frac{\mu\omega+x}{\sqrt{2}\sqrt{x}}\right) \tag{156}$$

with respect to x is

$$\frac{e^{\frac{(\mu\omega+x)^2}{2x}}\left(x(x+2)-\mu^2\omega^2\right)\operatorname{erfc}\left(\frac{\mu\omega+x}{\sqrt{2}\sqrt{x}}\right)}{2x} + \frac{\mu\omega-x}{\sqrt{2\pi}\sqrt{x}}\,. \tag{157}$$

This derivative is larger than zero, since

$$\begin{aligned}
&\frac{e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\left(\nu\tau(\nu\tau+2)-\mu^2\omega^2\right)\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)}{2\nu\tau} + \frac{\mu\omega-\nu\tau}{\sqrt{2\pi}\sqrt{\nu\tau}} \;> \qquad (158)\\
&\frac{0.4349\left(\nu\tau(\nu\tau+2)-\mu^2\omega^2\right)}{2\nu\tau} + \frac{\mu\omega-\nu\tau}{\sqrt{2\pi}\sqrt{\nu\tau}} \;> \\
&\frac{0.5\left(\nu\tau(\nu\tau+2)-\mu^2\omega^2\right)}{\sqrt{2\pi}\nu\tau} + \frac{\mu\omega-\nu\tau}{\sqrt{2\pi}\sqrt{\nu\tau}} \;= \\
&\frac{0.5\left(\nu\tau(\nu\tau+2)-\mu^2\omega^2\right)+\sqrt{\nu\tau}(\mu\omega-\nu\tau)}{\sqrt{2\pi}\nu\tau} \;=
\end{aligned}$$

$$\frac{-0.5\mu^2\omega^2 + \mu\omega\sqrt{\nu\tau} + 0.5(\nu\tau)^2 - \nu\tau\sqrt{\nu\tau} + \nu\tau}{\sqrt{2\pi}\nu\tau} =$$
$$\frac{-0.5\mu^2\omega^2 + \mu\omega\sqrt{\nu\tau} + \left(0.5\nu\tau - \sqrt{\nu\tau}\right)^2 + 0.25(\nu\tau)^2}{\sqrt{2\pi}\nu\tau} > 0 .$$

We explain this chain of inequalities:

- The first inequality follows by applying Lemma 23 which says that $e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$ is strictly monotonically decreasing. The minimal value that is larger than 0.4349 is taken on at the maximal values $\nu\tau = 1.5 \cdot 1.25$ and $\mu\omega = 0.1 \cdot 0.1$.
- The second inequality uses $\frac{1}{2}0.4349\sqrt{2\pi} = 0.545066 > 0.5$.
- The equalities are just algebraic reformulations.
- The last inequality follows from $-0.5\mu^2\omega^2 + \mu\omega\sqrt{\nu\tau} + 0.25(\nu\tau)^2 > 0.25(0.8 \cdot 0.8)^2 - 0.5 \cdot (0.1)^2(0.1)^2 - 0.1 \cdot 0.1 \cdot \sqrt{0.8 \cdot 0.8} = 0.09435 > 0$.

Therefore the function is increasing in $\nu\tau$.

Decreasing in $\mu\omega$ follows from decreasing of $e^{x^2}\operatorname{erfc}(x)$ according to Lemma 23. Positivity follows form the fact that erfc and the exponential function are positive and that $\nu\tau > 0$. □

Lemma 38 (Function $\nu\tau e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$)**.** *The function* $\nu\tau e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) > 0$ *is increasing in* $\nu\tau$ *and decreasing in* $\mu\omega$.

Proof. The derivative of the function

$$x e^{\frac{(\mu\omega+2x)^2}{2x}} \operatorname{erfc}\left(\frac{\mu\omega + 2x}{\sqrt{2}\sqrt{2x}}\right) \tag{159}$$

is

$$\frac{e^{\frac{(\mu\omega+2x)^2}{4x}}\left(\sqrt{\pi}e^{\frac{(\mu\omega+2x)^2}{4x}}\left(2x(2x+1) - \mu^2\omega^2\right)\operatorname{erfc}\left(\frac{\mu\omega+2x}{2\sqrt{x}}\right) + \sqrt{x}(\mu\omega - 2x)\right)}{2\sqrt{\pi}x} . \tag{160}$$

We only have to determine the sign of $\sqrt{\pi}e^{\frac{(\mu\omega+2x)^2}{4x}}\left(2x(2x+1) - \mu^2\omega^2\right)\operatorname{erfc}\left(\frac{\mu\omega+2x}{2\sqrt{x}}\right) + \sqrt{x}(\mu\omega - 2x)$ since all other factors are obviously larger than zero.

This derivative is larger than zero, since

$$\sqrt{\pi}e^{\frac{(\mu\omega+2\nu\tau)^2}{4\nu\tau}}\left(2\nu\tau(2\nu\tau+1) - \mu^2\omega^2\right)\operatorname{erfc}\left(\frac{\mu\omega + 2\nu\tau}{2\sqrt{\nu\tau}}\right) + \sqrt{\nu\tau}(\mu\omega - 2\nu\tau) > \tag{161}$$
$$0.463979\left(2\nu\tau(2\nu\tau+1) - \mu^2\omega^2\right) + \sqrt{\nu\tau}(\mu\omega - 2\nu\tau) =$$
$$-0.463979\mu^2\omega^2 + \mu\omega\sqrt{\nu\tau} + 1.85592(\nu\tau)^2 + 0.927958\nu\tau - 2\nu\tau\sqrt{\nu\tau} =$$
$$\mu\omega\left(\sqrt{\nu\tau} - 0.463979\mu\omega\right) + 0.85592(\nu\tau)^2 + \left(\nu\tau - \sqrt{\nu\tau}\right)^2 - 0.0720421\nu\tau > 0 .$$

We explain this chain of inequalities:

- The first inequality follows by applying Lemma 23 which says that $e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$ is strictly monotonically decreasing. The minimal value that is larger than 0.261772 is taken on at the maximal values $\nu\tau = 1.5 \cdot 1.25$ and $\mu\omega = 0.1 \cdot 0.1$. $0.261772\sqrt{\pi} > 0.463979$.
- The equalities are just algebraic reformulations.

- The last inequality follows from $\mu\omega\left(\sqrt{\nu\tau} - 0.463979\mu\omega\right) + 0.85592(\nu\tau)^2 - 0.0720421\nu\tau > 0.85592 \cdot (0.8 \cdot 0.8)^2 - 0.1 \cdot 0.1\left(\sqrt{1.5 \cdot 1.25} + 0.1 \cdot 0.1 \cdot 0.463979\right) - 0.0720421 \cdot 1.5 \cdot 1.25 > 0.201766$.

Therefore the function is increasing in $\nu\tau$.

Decreasing in $\mu\omega$ follows from decreasing of $e^{x^2}\operatorname{erfc}(x)$ according to Lemma 23. Positivity follows from the fact that erfc and the exponential function are positive and that $\nu\tau > 0$. $\square$

Lemma 39 (Bounds on the Derivatives)**.** *The following bounds on the absolute values of the derivatives of the Jacobian entries $\mathcal{J}_{11}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, $\mathcal{J}_{12}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, $\mathcal{J}_{21}(\mu,\omega,\nu,\tau,\lambda,\alpha)$, and $\mathcal{J}_{22}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ with respect to μ, ω, ν, and τ hold:*

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right| < 0.0031049101995398316 \tag{162}$$
$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| < 1.055872374194189$$
$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| < 0.03749149348255419$$

$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| < 0.031242911235461816$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| < 0.21232788238624354$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| < 0.2124377655377270$$

$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \mu}\right| < 0.02220441024325437$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \omega}\right| < 1.146955401845684$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \nu}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \tau}\right| < 0.17980135762932363$$

$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \mu}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \omega}\right| < 0.14983446469110305$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \nu}\right| < 1.805740052651535$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \tau}\right| < 2.396685907216327$$

Proof. For each derivative we compute a lower and an upper bound and take the maximum of the absolute value. A lower bound is determined by minimizing the single terms of the functions that represents the derivative. An upper bound is determined by maximizing the single terms of the functions that represent the derivative. Terms can be combined to larger terms for which the maximum and the minimum must be known. We apply many previous lemmata which state properties of functions representing single or combined terms. The more terms are combined, the tighter the bounds can be made.

Next we go through all the derivatives, where we use Lemma 25, Lemma 26, Lemma 27, Lemma 28, Lemma 29, Lemma 30, Lemma 21, and Lemma 23 without citing. Furthermore, we use the bounds on the simple expressions t_{11},t_{22}, ..., and T_4 as defined the aforementioned lemmata:

- $\frac{\partial \mathcal{J}_{11}}{\partial \mu}$

  We use Lemma 31 and consider the expression $\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - \frac{\sqrt{\frac{2}{\pi}}(\alpha-1)}{\sqrt{\nu\tau}}$ in brackets. An upper bound on the maximum of is

$$\alpha_{01} e^{t_1^2} \operatorname{erfc}(t_1) - \frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)}{\sqrt{T_{22}}} \;=\; 0.591017 \,. \tag{163}$$

  A lower bound on the minimum is

$$\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) - \frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)}{\sqrt{t_{22}}} \;=\; 0.056318 \,. \tag{164}$$

  Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}\omega_{\max}^2 e^{t_4}\left(\alpha_{01} e^{t_1^2} \operatorname{erfc}(t_1) - \frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)}{\sqrt{T_{22}}}\right) \;=\; 0.0031049101995398316 \,. \tag{165}$$

- $\frac{\partial \mathcal{J}_{11}}{\partial \omega}$

  We use Lemma 31 and consider the expression $\frac{\sqrt{\frac{2}{\pi}}(\alpha-1)\mu\omega}{\sqrt{\nu\tau}} - \alpha(\mu\omega+1)e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$ in brackets.

  An upper bound on the maximum is

$$\frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)T_{11}}{\sqrt{t_{22}}} - \alpha_{01}(t_{11}+1)e^{T_1^2} \operatorname{erfc}(T_1) \;=\; -0.713808 \,. \tag{166}$$

  A lower bound on the minimum is

$$\frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)t_{11}}{\sqrt{t_{22}}} - \alpha_{01}(T_{11}+1)e^{t_1^2} \operatorname{erfc}(t_1) \;=\; -0.99987 \,. \tag{167}$$

  This term is subtracted, and $2-\operatorname{erfc}(x)>0$, therefore we have to use the minimum and the maximum for the argument of erfc.

  Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}\left(-e^{t_4}\left(\frac{\sqrt{\frac{2}{\pi}}(\alpha_{01}-1)t_{11}}{\sqrt{t_{22}}} - \alpha_{01}(T_{11}+1)e^{t_1^2} \operatorname{erfc}(t_1)\right) - \operatorname{erfc}(T_3)+2\right) \;= \tag{168}$$

  1.055872374194189 .

- $\frac{\partial \mathcal{J}_{11}}{\partial \nu}$
  We consider the term in brackets
$$\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha-1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{\alpha}{\sqrt{\nu\tau}}\right) . \tag{169}$$
  We apply Lemma 33 for the first sub-term. An upper bound on the maximum is
$$\alpha_{01} e^{t_1^2} \operatorname{erfc}(t_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)T_{11}}{T_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{T_{22}}}\right) = 0.0104167 . \tag{170}$$
  A lower bound on the minimum is
$$\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)t_{11}}{t_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{t_{22}}}\right) = -0.95153 . \tag{171}$$
  Thus, an upper bound on the maximal absolute value is
$$-\frac{1}{4}\lambda_{01}\tau_{\max}\omega_{\max}e^{t_4}\left(\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)t_{11}}{t_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{t_{22}}}\right)\right) = \tag{172}$$
$$0.031242911235461816 .$$

- $\frac{\partial \mathcal{J}_{11}}{\partial \tau}$
  We use the results of item $\frac{\partial \mathcal{J}_{11}}{\partial \nu}$ were the brackets are only differently scaled. Thus, an upper bound on the maximal absolute value is
$$-\frac{1}{4}\lambda_{01}\nu_{\max}\omega_{\max}e^{t_4}\left(\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)t_{11}}{t_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{t_{22}}}\right)\right) = \tag{173}$$
$$0.03749149348255419 .$$

- $\frac{\partial \mathcal{J}_{12}}{\partial \mu}$
  Since $\frac{\partial \mathcal{J}_{12}}{\partial \mu} = \frac{\partial \mathcal{J}_{11}}{\partial \nu}$, an upper bound on the maximal absolute value is
$$-\frac{1}{4}\lambda_{01}\tau_{\max}\omega_{\max}e^{t_4}\left(\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)t_{11}}{t_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{t_{22}}}\right)\right) = \tag{174}$$
$$0.031242911235461816 .$$

- $\frac{\partial \mathcal{J}_{12}}{\partial \omega}$
  We use the results of item $\frac{\partial \mathcal{J}_{11}}{\partial \nu}$ were the brackets are only differently scaled. Thus, an upper bound on the maximal absolute value is
$$-\frac{1}{4}\lambda_{01}\mu_{\max}\tau_{\max}e^{t_4}\left(\alpha_{01} e^{T_1^2} \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\left(\frac{(\alpha_{01}-1)t_{11}}{t_{22}^{3/2}} - \frac{\alpha_{01}}{\sqrt{t_{22}}}\right)\right) = \tag{175}$$
$$0.031242911235461816 .$$

- $\frac{\partial \mathcal{J}_{12}}{\partial \nu}$
  For the second term in brackets, we see that $\alpha_{01}\tau_{\min}^2 e^{T_1^2} \operatorname{erfc}(T_1) = 0.465793$ and $\alpha_{01}\tau_{\max}^2 e^{t_1^2} \operatorname{erfc}(t_1) = 1.53644$.
  We now check different values for
$$\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha-1)\mu^2\omega^2}{\nu^{5/2}\sqrt{\tau}} + \frac{\sqrt{\tau}(\alpha+\alpha\mu\omega-1)}{\nu^{3/2}} - \frac{\alpha\tau^{3/2}}{\sqrt{\nu}}\right) , \tag{176}$$

where we maximize or minimize all single terms.

A lower bound on the minimum of this expression is

$$\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha_{01}-1)\mu_{\max}^2\omega_{\max}^2}{\nu_{\min}^{5/2}\sqrt{\tau_{\min}}}+\frac{\sqrt{\tau_{\min}}(\alpha_{01}+\alpha_{01}t_{11}-1)}{\nu_{\max}^{3/2}}-\frac{\alpha_{01}\tau_{\max}^{3/2}}{\sqrt{\nu_{\min}}}\right) = -1.83112\,. \tag{177}$$

An upper bound on the maximum of this expression is

$$\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha_{01}-1)\mu_{\min}^2\omega_{\min}^2}{\nu_{\max}^{5/2}\sqrt{\tau_{\max}}}+\frac{\sqrt{\tau_{\max}}(\alpha_{01}+\alpha_{01}T_{11}-1)}{\nu_{\min}^{3/2}}-\frac{\alpha_{01}\tau_{\min}^{3/2}}{\sqrt{\nu_{\max}}}\right) = 0.0802158\,. \tag{178}$$

An upper bound on the maximum is

$$\frac{1}{8}\lambda_{01}e^{t_4}\left(\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha_{01}-1)\mu_{\min}^2\omega_{\min}^2}{\nu_{\max}^{5/2}\sqrt{\tau_{\max}}}-\frac{\alpha_{01}\tau_{\min}^{3/2}}{\sqrt{\nu_{\max}}}+\frac{\sqrt{\tau_{\max}}(\alpha_{01}+\alpha_{01}T_{11}-1)}{\nu_{\min}^{3/2}}\right)+\alpha_{01}\tau_{\max}^2e^{t_1^2}\operatorname{erfc}(t_1)\right) = 0.212328\,. \tag{179}$$

A lower bound on the minimum is

$$\frac{1}{8}\lambda_{01}e^{t_4}\left(\alpha_{01}\tau_{\min}^2e^{T_1^2}\operatorname{erfc}(T_1)+\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha_{01}-1)\mu_{\max}^2\omega_{\max}^2}{\nu_{\min}^{5/2}\sqrt{\tau_{\min}}}+\frac{\sqrt{\tau_{\min}}(\alpha_{01}+\alpha_{01}t_{11}-1)}{\nu_{\max}^{3/2}}-\frac{\alpha_{01}\tau_{\max}^{3/2}}{\sqrt{\nu_{\min}}}\right)\right) = -0.179318\,. \tag{180}$$

Thus, an upper bound on the maximal absolute value is

$$\frac{1}{8}\lambda_{01}e^{t_4}\left(\sqrt{\frac{2}{\pi}}\left(\frac{(-1)(\alpha_{01}-1)\mu_{\min}^2\omega_{\min}^2}{\nu_{\max}^{5/2}\sqrt{\tau_{\max}}}-\frac{\alpha_{01}\tau_{\min}^{3/2}}{\sqrt{\nu_{\max}}}+\frac{\sqrt{\tau_{\max}}(\alpha_{01}+\alpha_{01}T_{11}-1)}{\nu_{\min}^{3/2}}\right)+\alpha_{01}\tau_{\max}^2e^{t_1^2}\operatorname{erfc}(t_1)\right) = 0.21232788238624354\,. \tag{181}$$

- $\frac{\partial \mathcal{J}_{12}}{\partial \tau}$

We use Lemma 34 to obtain an upper bound on the maximum of the expression of the lemma:

$$\sqrt{\frac{2}{\pi}}\left(\frac{0.1^2\cdot 0.1^2(-1)(\alpha_{01}-1)}{(0.8\cdot 0.8)^{3/2}}-\sqrt{0.8\cdot 0.8}\alpha_{01}+\frac{(0.1\cdot 0.1)\alpha_{01}-\alpha_{01}+1}{\sqrt{0.8\cdot 0.8}}\right) = -1.72296\,. \tag{182}$$

We use Lemma 34 to obtain an lower bound on the minimum of the expression of the lemma:

$$\sqrt{\frac{2}{\pi}}\left(\frac{0.1^2\cdot 0.1^2(-1)(\alpha_{01}-1)}{(1.5\cdot 1.25)^{3/2}}-\sqrt{1.5\cdot 1.25}\alpha_{01}+\frac{(-0.1\cdot 0.1)\alpha_{01}-\alpha_{01}+1}{\sqrt{1.5\cdot 1.25}}\right) = -2.2302\,. \tag{183}$$

Next we apply Lemma 37 for the expression $\nu\tau e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$. We use Lemma 37 to obtain an upper bound on the maximum of this expression:

$$1.5\cdot 1.25e^{\frac{(1.5\cdot 1.25-0.1\cdot 0.1)^2}{2\cdot 1.5\cdot 1.25}}\alpha_{01}\operatorname{erfc}\left(\frac{1.5\cdot 1.25-0.1\cdot 0.1}{\sqrt{2}\sqrt{1.5\cdot 1.25}}\right) = 1.37381\,. \tag{184}$$

We use Lemma 37 to obtain an lower bound on the minimum of this expression:

$$0.8 \cdot 0.8 e^{\frac{(0.8 \cdot 0.8 + 0.1 \cdot 0.1)^2}{2 \cdot 0.8 \cdot 0.8}} \alpha_{01} \operatorname{erfc}\left(\frac{0.8 \cdot 0.8 + 0.1 \cdot 0.1}{\sqrt{2}\sqrt{0.8 \cdot 0.8}}\right) = 0.620462 . \tag{185}$$

Next we apply Lemma 23 for $2\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$. An upper bound on this expression is

$$2 e^{\frac{(0.8 \cdot 0.8 - 0.1 \cdot 0.1)^2}{20.8 \cdot 0.8}} \alpha_{01} \operatorname{erfc}\left(\frac{0.8 \cdot 0.8 - 0.1 \cdot 0.1}{\sqrt{2}\sqrt{0.8 \cdot 0.8}}\right) = 1.96664 . \tag{186}$$

A lower bound on this expression is

$$2 e^{\frac{(1.5 \cdot 1.25 + 0.1 \cdot 0.1)^2}{2 \cdot 1.5 \cdot 1.25}} \alpha_{01} \operatorname{erfc}\left(\frac{1.5 \cdot 1.25 + 0.1 \cdot 0.1}{\sqrt{2}\sqrt{1.5 \cdot 1.25}}\right) = 1.4556 . \tag{187}$$

The sum of the minimal values of the terms is $-2.23019 + 0.62046 + 1.45560 = -0.154133$. The sum of the maximal values of the terms is $-1.72295 + 1.37380 + 1.96664 = 1.61749$.

Thus, an upper bound on the maximal absolute value is

$$\frac{1}{8}\lambda_{01} e^{t_4}\left(\alpha_{01} T_{22} e^{\frac{(t_{11}+T_{22})^2}{2T_{22}}} \operatorname{erfc}\left(\frac{t_{11}+T_{22}}{\sqrt{2}\sqrt{T_{22}}}\right) + \right. \tag{188}$$
$$2\alpha_{01} e^{t_1^2} \operatorname{erfc}(t_1) + \sqrt{\frac{2}{\pi}}\left(-\frac{(\alpha_{01}-1)T_{11}^2}{t_{22}^{3/2}} + \frac{-\alpha_{01}+\alpha_{01}T_{11}+1}{\sqrt{t_{22}}} - \right.$$
$$\left.\left.\alpha_{01}\sqrt{t_{22}}\right)\right) = 0.2124377655377270 .$$

- $\frac{\partial \mathcal{J}_{21}}{\partial \mu}$

  An upper bound on the maximum is

$$\lambda_{01}^2 \omega_{\max}^2 \left(\alpha_{01}^2 e^{T_1^2}\left(-e^{-T_4}\right) \operatorname{erfc}(T_1) + 2\alpha_{01}^2 e^{t_2^2} e^{t_4} \operatorname{erfc}(t_2) - \operatorname{erfc}(T_3) + 2\right) = \tag{189}$$
$$0.0222044 .$$

  A upper bound on the absolute minimum is

$$\lambda_{01}^2 \omega_{\max}^2 \left(\alpha_{01}^2 e^{t_1^2}\left(-e^{-t_4}\right) \operatorname{erfc}(t_1) + 2\alpha_{01}^2 e^{T_2^2} e^{T_4} \operatorname{erfc}(T_2) - \operatorname{erfc}(t_3) + 2\right) = \tag{190}$$
$$0.00894889 .$$

  Thus, an upper bound on the maximal absolute value is

$$\lambda_{01}^2 \omega_{\max}^2 \left(\alpha_{01}^2 e^{T_1^2}\left(-e^{-T_4}\right) \operatorname{erfc}(T_1) + 2\alpha_{01}^2 e^{t_2^2} e^{t_4} \operatorname{erfc}(t_2) - \operatorname{erfc}(T_3) + 2\right) = \tag{191}$$
$$0.02220441024325437 .$$

- $\frac{\partial \mathcal{J}_{21}}{\partial \omega}$

  An upper bound on the maximum is

$$\lambda_{01}^2 \left(\alpha_{01}^2 (2T_{11}+1) e^{t_2^2} e^{-t_4} \operatorname{erfc}(t_2) + 2T_{11}(2 - \operatorname{erfc}(T_3)) + \right. \tag{192}$$
$$\left.\alpha_{01}^2 (t_{11}+1) e^{T_1^2}\left(-e^{-T_4}\right) \operatorname{erfc}(T_1) + \sqrt{\frac{2}{\pi}}\sqrt{T_{22}} e^{-t_4}\right) = 1.14696 .$$

  A lower bound on the minimum is

$$\lambda_{01}^2 \left(\alpha_{01}^2 (T_{11}+1) e^{t_1^2}\left(-e^{-t_4}\right) \operatorname{erfc}(t_1) + \right. \tag{193}$$
$$\alpha_{01}^2 (2t_{11}+1) e^{T_2^2} e^{-T_4} \operatorname{erfc}(T_2) + 2t_{11}(2 - \operatorname{erfc}(T_3)) +$$

$$\left.\sqrt{\frac{2}{\pi}}\sqrt{t_{22}}e^{-T_4}\right) \;=\; -0.359403\;.$$

Thus, an upper bound on the maximal absolute value is

$$\lambda_{01}^2\left(\alpha_{01}^2(2T_{11}+1)e^{t_2^2}e^{-t_4}\operatorname{erfc}(t_2)+2T_{11}(2-\operatorname{erfc}(T_3))+\right. \tag{194}$$
$$\left.\alpha_{01}^2(t_{11}+1)e^{T_1^2}\left(-e^{-T_4}\right)\operatorname{erfc}(T_1)+\sqrt{\frac{2}{\pi}}\sqrt{T_{22}}e^{-t_4}\right) \;=\; 1.146955401845684\;.$$

- $\frac{\partial \mathcal{J}_{21}}{\partial \nu}$

  An upper bound on the maximum is

$$\frac{1}{2}\lambda_{01}^2\tau_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{T_1^2}\right)\operatorname{erfc}(T_1)+4\alpha_{01}^2e^{t_2^2}\operatorname{erfc}(t_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{T_{22}}}\right) \;= \tag{195}$$

  $0.149834\;.$

  A lower bound on the minimum is

$$\frac{1}{2}\lambda_{01}^2\tau_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{t_1^2}\right)\operatorname{erfc}(t_1)+4\alpha_{01}^2e^{T_2^2}\operatorname{erfc}(T_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{t_{22}}}\right) \;= \tag{196}$$

  $-\;0.0351035\;.$

  Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}^2\tau_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{T_1^2}\right)\operatorname{erfc}(T_1)+4\alpha_{01}^2e^{t_2^2}\operatorname{erfc}(t_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{T_{22}}}\right) \;= \tag{197}$$

  $0.14983446469110305\;.$

- $\frac{\partial \mathcal{J}_{21}}{\partial \tau}$

  An upper bound on the maximum is

$$\frac{1}{2}\lambda_{01}^2\nu_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{T_1^2}\right)\operatorname{erfc}(T_1)+4\alpha_{01}^2e^{t_2^2}\operatorname{erfc}(t_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{T_{22}}}\right) \;= \tag{198}$$

  $0.179801\;.$

  A lower bound on the minimum is

$$\frac{1}{2}\lambda_{01}^2\nu_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{t_1^2}\right)\operatorname{erfc}(t_1)+4\alpha_{01}^2e^{T_2^2}\operatorname{erfc}(T_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{t_{22}}}\right) \;= \tag{199}$$

  $-\;0.0421242\;.$

  Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}^2\nu_{\max}\omega_{\max}e^{-t_4}\left(\alpha_{01}^2\left(-e^{T_1^2}\right)\operatorname{erfc}(T_1)+4\alpha_{01}^2e^{t_2^2}\operatorname{erfc}(t_2)\;+\;\frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2-1\right)}{\sqrt{T_{22}}}\right) \;= \tag{200}$$

  $0.17980135762932363\;.$

- $\frac{\partial \mathcal{J}_{22}}{\partial \mu}$

  We use the fact that $\frac{\partial \mathcal{J}_{22}}{\partial \mu} = \frac{\partial \mathcal{J}_{21}}{\partial \nu}$. Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}^2 \tau_{\max}\omega_{\max} e^{-t_4} \left(\alpha_{01}^2 \left(-e^{T_1^2}\right) \operatorname{erfc}(T_1) + 4\alpha_{01}^2 e^{t_2^2} \operatorname{erfc}(t_2) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2 - 1\right)}{\sqrt{T_{22}}}\right) = \tag{201}$$

  0.14983446469110305 .

- $\frac{\partial \mathcal{J}_{22}}{\partial \omega}$

  An upper bound on the maximum is

$$\frac{1}{2}\lambda_{01}^2 \mu_{\max}\tau_{\max} e^{-t_4} \left(\alpha_{01}^2 \left(-e^{T_1^2}\right) \operatorname{erfc}(T_1) + 4\alpha_{01}^2 e^{t_2^2} \operatorname{erfc}(t_2) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2 - 1\right)}{\sqrt{T_{22}}}\right) = \tag{202}$$

  0.149834 .

  A lower bound on the minimum is

$$\frac{1}{2}\lambda_{01}^2 \mu_{\max}\tau_{\max} e^{-t_4} \left(\alpha_{01}^2 \left(-e^{t_1^2}\right) \operatorname{erfc}(t_1) + 4\alpha_{01}^2 e^{T_2^2} \operatorname{erfc}(T_2) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2 - 1\right)}{\sqrt{t_{22}}}\right) = \tag{203}$$

  $-\,0.0351035$.

  Thus, an upper bound on the maximal absolute value is

$$\frac{1}{2}\lambda_{01}^2 \mu_{\max}\tau_{\max} e^{-t_4} \left(\alpha_{01}^2 \left(-e^{T_1^2}\right) \operatorname{erfc}(T_1) + 4\alpha_{01}^2 e^{t_2^2} \operatorname{erfc}(t_2) + \frac{\sqrt{\frac{2}{\pi}}(-1)\left(\alpha_{01}^2 - 1\right)}{\sqrt{T_{22}}}\right) = \tag{204}$$

  0.14983446469110305 .

- $\frac{\partial \mathcal{J}_{22}}{\partial \nu}$

  We apply Lemma 35 to the expression $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{(\nu\tau)^{3/2}} - \frac{3\alpha^2}{\sqrt{\nu\tau}}\right)$. Using Lemma 35, an upper bound on the maximum is

$$\frac{1}{4}\lambda_{01}^2 \tau_{\max}^2 e^{-t_4} \left(\alpha_{01}^2 \left(-e^{T_1^2}\right) \operatorname{erfc}(T_1) + 8\alpha_{01}^2 e^{t_2^2} \operatorname{erfc}(t_2) + \right. \tag{205}$$
$$\left.\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha_{01}^2 - 1\right) T_{11}}{T_{22}^{3/2}} - \frac{3\alpha_{01}^2}{\sqrt{T_{22}}}\right)\right) = 1.19441 \, .$$

  Using Lemma 35, a lower bound on the minimum is

$$\frac{1}{4}\lambda_{01}^2 \tau_{\max}^2 e^{-t_4} \left(\alpha_{01}^2 \left(-e^{t_1^2}\right) \operatorname{erfc}(t_1) + 8\alpha_{01}^2 e^{T_2^2} \operatorname{erfc}(T_2) + \right. \tag{206}$$
$$\left.\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha_{01}^2 - 1\right) t_{11}}{t_{22}^{3/2}} - \frac{3\alpha_{01}^2}{\sqrt{t_{22}}}\right)\right) = -1.80574 \, .$$

  Thus, an upper bound on the maximal absolute value is

$$-\frac{1}{4}\lambda_{01}^2 \tau_{\max}^2 e^{-t_4} \left(\alpha_{01}^2 \left(-e^{t_1^2}\right) \operatorname{erfc}(t_1) + 8\alpha_{01}^2 e^{T_2^2} \operatorname{erfc}(T_2) + \right. \tag{207}$$
$$\left.\sqrt{\frac{2}{\pi}}\left(\frac{\left(\alpha_{01}^2 - 1\right) t_{11}}{t_{22}^{3/2}} - \frac{3\alpha_{01}^2}{\sqrt{t_{22}}}\right)\right) = 1.805740052651535 \, .$$

- $\frac{\partial \mathcal{J}_{22}}{\partial \tau}$

  We apply Lemma 36 to the expression $\sqrt{\frac{2}{\pi}}\left(\frac{(\alpha^2-1)\mu\omega}{\sqrt{\nu\tau}} - 3\alpha^2\sqrt{\nu\tau}\right)$.

  We apply Lemma 37 to the expression $\nu\tau e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$. We apply Lemma 38 to the expression $\nu\tau e^{\frac{(\mu\omega+2\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$.

  We combine the results of these lemmata to obtain an upper bound on the maximum:

$$
\begin{aligned}
&\frac{1}{4}\lambda_{01}^2 \left(-\alpha_{01}^2 t_{22} e^{-T_4} e^{\frac{(T_{11}+t_{22})^2}{2t_{22}}} \operatorname{erfc}\left(\frac{T_{11}+t_{22}}{\sqrt{2}\sqrt{t_{22}}}\right) + \right. \\
&8\alpha_{01}^2 T_{22} e^{-t_4} e^{\frac{(t_{11}+2T_{22})^2}{2T_{22}}} \operatorname{erfc}\left(\frac{t_{11}+2T_{22}}{\sqrt{2}\sqrt{T_{22}}}\right) - \\
&2\alpha_{01}^2 e^{T_1^2} e^{-T_4} \operatorname{erfc}(T_1) + 4\alpha_{01}^2 e^{t_2^2} e^{-t_4} \operatorname{erfc}(t_2) + 2(2-\operatorname{erfc}(T_3)) + \\
&\left.\sqrt{\frac{2}{\pi}} e^{-T_4}\left(\frac{\left(\alpha_{01}^2-1\right) T_{11}}{\sqrt{t_{22}}} - 3\alpha_{01}^2\sqrt{t_{22}}\right)\right) \;=\; 2.39669\,.
\end{aligned} \tag{208}
$$

  We combine the results of these lemmata to obtain an lower bound on the minimum:

$$
\begin{aligned}
&\frac{1}{4}\lambda_{01}^2 \left(8\alpha_{01}^2 t_{22} e^{-T_4} e^{\frac{(T_{11}+2t_{22})^2}{2t_{22}}} \operatorname{erfc}\left(\frac{T_{11}+2t_{22}}{\sqrt{2}\sqrt{t_{22}}}\right) + \right. \\
&\alpha_{01}^2 T_{22} e^{-t_4} e^{\frac{(t_{11}+T_{22})^2}{2T_{22}}} \operatorname{erfc}\left(\frac{t_{11}+T_{22}}{\sqrt{2}\sqrt{T_{22}}}\right) - \\
&2\alpha_{01}^2 e^{t_1^2} e^{-t_4} \operatorname{erfc}(t_1) + 4\alpha_{01}^2 e^{T_2^2} e^{-T_4} \operatorname{erfc}(T_2) + \\
&\left. 2(2-\operatorname{erfc}(t_3)) + \sqrt{\frac{2}{\pi}} e^{-t_4}\left(\frac{\left(\alpha_{01}^2-1\right) t_{11}}{\sqrt{T_{22}}} - 3\alpha_{01}^2\sqrt{T_{22}}\right)\right) \;=\; -1.17154\,.
\end{aligned} \tag{209}
$$

  Thus, an upper bound on the maximal absolute value is

$$
\begin{aligned}
&\frac{1}{4}\lambda_{01}^2 \left(-\alpha_{01}^2 t_{22} e^{-T_4} e^{\frac{(T_{11}+t_{22})^2}{2t_{22}}} \operatorname{erfc}\left(\frac{T_{11}+t_{22}}{\sqrt{2}\sqrt{t_{22}}}\right) + \right. \\
&8\alpha_{01}^2 T_{22} e^{-t_4} e^{\frac{(t_{11}+2T_{22})^2}{2T_{22}}} \operatorname{erfc}\left(\frac{t_{11}+2T_{22}}{\sqrt{2}\sqrt{T_{22}}}\right) - \\
&2\alpha_{01}^2 e^{T_1^2} e^{-T_4} \operatorname{erfc}(T_1) + 4\alpha_{01}^2 e^{t_2^2} e^{-t_4} \operatorname{erfc}(t_2) + 2(2-\operatorname{erfc}(T_3)) + \\
&\left.\sqrt{\frac{2}{\pi}} e^{-T_4}\left(\frac{\left(\alpha_{01}^2-1\right) T_{11}}{\sqrt{t_{22}}} - 3\alpha_{01}^2\sqrt{t_{22}}\right)\right) \;=\; 2.396685907216327\,.
\end{aligned} \tag{210}
$$

□

Lemma 40 (Derivatives of the Mapping)**.** *We assume $\alpha = \alpha_{01}$ and $\lambda = \lambda_{01}$. We restrict the range of the variables to the domain $\mu \in [-0.1, 0.1]$, $\omega \in [-0.1, 0.1]$, $\nu \in [0.8, 1.5]$, and $\tau \in [0.8, 1.25]$.*

The derivative $\frac{\partial}{\partial \mu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ has the sign of ω.

The derivative $\frac{\partial}{\partial \nu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ is positive.

The derivative $\frac{\partial}{\partial \mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ has the sign of ω.

The derivative $\frac{\partial}{\partial \nu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ is positive.

Proof.

- $\frac{\partial}{\partial \mu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$

  $(2 - \operatorname{erfc}(x) > 0$ according to Lemma 21 and $e^{x^2}\operatorname{erfc}(x)$ is also larger than zero according to Lemma 23. Consequently, has $\frac{\partial}{\partial \mu}\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ the sign of ω.

- $\frac{\partial}{\partial \nu}\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha)$

  Lemma 23 says $e^{x^2}\operatorname{erfc}(x)$ is decreasing in $\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}$. The first term (negative) is increasing in $\nu\tau$ since it is proportional to minus one over the squared root of $\nu\tau$.

  We obtain a lower bound by setting $\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}} = \frac{1.5\cdot1.25+0.1\cdot0.1}{\sqrt{2}\sqrt{1.5\cdot1.25}}$ for the $e^{x^2}\operatorname{erfc}(x)$ term. The term in brackets is larger than $e^{\left(\frac{1.5\cdot1.25+0.1\cdot0.1}{\sqrt{2}\sqrt{1.5\cdot1.25}}\right)^2}\alpha_{01}\operatorname{erfc}\left(\frac{1.5\cdot1.25+0.1\cdot0.1}{\sqrt{2}\sqrt{1.5\cdot1.25}}\right) - \sqrt{\frac{2}{\pi 0.8\cdot0.8}}(\alpha_{01}-1) = 0.056$ Consequently, the function is larger than zero.

- $\frac{\partial}{\partial \mu}\tilde{\xi}(\mu, \omega, \nu, \tau, \lambda, \alpha)$

  We consider the sub-function

$$\sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} - \alpha^2\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right). \tag{211}$$

  We set $x = \nu\tau$ and $y = \mu\omega$ and obtain

$$\sqrt{\frac{2}{\pi}}\sqrt{x} - \alpha^2\left(e^{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right). \tag{212}$$

  The derivative of this sub-function with respect to y is

$$\frac{\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}(2x+y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\frac{(x+y)^2}{2x}}(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right)}{x} = \tag{213}$$

$$\frac{\sqrt{2}\alpha^2\sqrt{x}\left(\frac{e^{\frac{(2x+y)^2}{2x}}(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}} - \frac{e^{\frac{(x+y)^2}{2x}}(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)}{x} > 0.$$

  The inequality follows from Lemma 24, which states that $ze^{z^2}\operatorname{erfc}(z)$ is monotonically increasing in z. Therefore the sub-function is increasing in y.

  The derivative of this sub-function with respect to x is

$$\frac{\sqrt{\pi}\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right) - \sqrt{2}\left(\alpha^2-1\right)x^{3/2}}{2\sqrt{\pi}x^2}. \tag{214}$$

  The sub-function is increasing in x, since the derivative is larger than zero:

$$\frac{\sqrt{\pi}\alpha^2\left(e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)\right) - \sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} \geqslant \tag{215}$$

$$\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2+2}\right)} - \frac{(x-y)(x+y)2}{\sqrt{\pi}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2+\frac{4}{\pi}}\right)}\right) - \sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} =$$

$$\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2\left(\sqrt{2}\sqrt{x}\right)}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+4x}\right)} - \frac{(x-y)(x+y)2\left(\sqrt{2}\sqrt{x}\right)}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+\frac{8x}{\pi}}\right)}\right) - \sqrt{2}x^{3/2}\left(\alpha^2-1\right)}{2\sqrt{\pi}x^2} =$$

$$\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+4x}\right)} - \frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+\frac{8x}{\pi}}\right)}\right) - x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} >$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y)^2+2(2x+y)+1}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y)^2+0.782\cdot 2(x+y)+0.782^2}\right)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}\left(2x+y+\sqrt{(2x+y+1)^2}\right)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}\left(x+y+\sqrt{(x+y+0.782)^2}\right)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2x-y)(2x+y)2}{\sqrt{\pi}(2(2x+y)+1)}-\frac{(x-y)(x+y)2}{\sqrt{\pi}(2(x+y)+0.782)}\right)-x\left(\alpha^2-1\right)}{\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(\frac{(2(x+y)+0.782)(2x-y)(2x+y)2}{\sqrt{\pi}}-\frac{(x-y)(x+y)(2(2x+y)+1)2}{\sqrt{\pi}}\right)}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}x^{3/2}} +
$$

$$
\frac{\sqrt{\pi}\alpha^2\left(-x\left(\alpha^2-1\right)(2(2x+y)+1)(2(x+y)+0.782)\right)}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{8x^3+(12y+2.68657)x^2+(y(4y-6.41452)-1.40745)x+1.22072y^2}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}x^{3/2}} >
$$

$$
\frac{8x^3+(2.68657-120.01)x^2+(0.01(-6.41452-40.01)-1.40745)x+1.22072(0.0)^2}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}x^{3/2}} =
$$

$$
\frac{8x^2+2.56657x-1.472}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}\sqrt{x}} =
$$

$$
\frac{8x^2+2.56657x-1.472}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}\sqrt{x}} =
$$

$$
\frac{8(x+0.618374)(x-0.297553)}{(2(2x+y)+1)(2(x+y)+0.782)\sqrt{2}\sqrt{\pi}\sqrt{x}} > 0 .
$$

We explain this chain of inequalities:

- First inequality: We applied Lemma 22 two times.
- Equalities factor out $\sqrt{2}\sqrt{x}$ and reformulate.
- Second inequality part 1: we applied

$$
0 < 2y \Longrightarrow (2x+y)^2+4x+1 < (2x+y)^2+2(2x+y)+1 = (2x+y+1)^2 . \tag{216}
$$

- Second inequality part 2: we show that for $a = \frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}}-13\right)$ following holds: $\frac{8x}{\pi}-\left(a^2+2a(x+y)\right) \geqslant 0$. We have $\frac{\partial}{\partial x}\frac{8x}{\pi}-\left(a^2+2a(x+y)\right) = \frac{8}{\pi}-2a > 0$ and $\frac{\partial}{\partial y}\frac{8x}{\pi}-\left(a^2+2a(x+y)\right) = -2a > 0$. Therefore the minimum is at border for minimal x and maximal y:

$$
\frac{8\cdot 0.64}{\pi}-\left(\frac{2}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}}-13\right)(0.64+0.01)+\left(\frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}}-13\right)\right)^2\right) = 0 . \tag{217}
$$

  Thus

$$
\frac{8x}{\pi} \geqslant a^2+2a(x+y) . \tag{218}
$$

  for $a = \frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}}-13\right) > 0.782$.
- Equalities only solve square root and factor out the resulting terms $(2(2x+y)+1)$ and $(2(x+y)+0.782)$.
- We set $\alpha = \alpha_{01}$ and multiplied out. Thereafter we also factored out x in the numerator. Finally a quadratic equations was solved.

The sub-function has its minimal value for minimal x and minimal y $x = \nu\tau = 0.8 \cdot 0.8 = 0.64$ and $y = \mu\omega = -0.1 \cdot 0.1 = -0.01$. We further minimize the function

$$\mu\omega e^{\frac{\mu^2\omega^2}{2\nu\tau}} \left(2 - \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) > -0.01 e^{\frac{0.01^2}{20.64}} \left(2 - \mathrm{erfc}\left(\frac{0.01}{\sqrt{2}\sqrt{0.64}}\right)\right) . \tag{219}$$

We compute the minimum of the term in brackets of $\frac{\partial}{\partial\mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$:

$$\begin{aligned}
&\mu\omega e^{\frac{\mu^2\omega^2}{2\nu\tau}} \left(2 - \mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) + \\
&\alpha_{01}^2 \left(-\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \mathrm{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \mathrm{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right)\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau} > \\
&\alpha_{01}^2 \left(-\left(e^{\left(\frac{0.64-0.01}{\sqrt{2}\sqrt{0.64}}\right)^2} \mathrm{erfc}\left(\frac{0.64-0.01}{\sqrt{2}\sqrt{0.64}}\right) - e^{\left(\frac{20.64-0.01}{\sqrt{2}\sqrt{0.64}}\right)^2} \mathrm{erfc}\left(\frac{2\cdot 0.64-0.01}{\sqrt{2}\sqrt{0.64}}\right)\right)\right) - \\
&0.01 e^{\frac{0.01^2}{20.64}} \left(2 - \mathrm{erfc}\left(\frac{0.01}{\sqrt{2}\sqrt{0.64}}\right)\right) + \sqrt{0.64}\sqrt{\frac{2}{\pi}} \;=\; 0.0923765 .
\end{aligned} \tag{220}$$

Therefore the term in brackets is larger than zero.

Thus, $\frac{\partial}{\partial\mu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ has the sign of ω.

- $\frac{\partial}{\partial\nu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$

  We look at the sub-term

$$2e^{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2} \mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2} \mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) . \tag{221}$$

  We obtain a chain of inequalities:

$$\begin{aligned}
&2e^{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2} \mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2} \mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) > \\
&\frac{2\cdot 2}{\sqrt{\pi}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}} + \sqrt{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2 + 2}\right)} - \frac{2}{\sqrt{\pi}\left(\frac{x+y}{\sqrt{2}\sqrt{x}} + \sqrt{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2 + \frac{4}{\pi}}\right)} = \\
&\frac{2\sqrt{2}\sqrt{x}\left(\frac{2}{\sqrt{(2x+y)^2+4x}+2x+y} - \frac{1}{\sqrt{(x+y)^2+\frac{8x}{\pi}}+x+y}\right)}{\sqrt{\pi}} > \\
&\frac{2\sqrt{2}\sqrt{x}\left(\frac{2}{\sqrt{(2x+y)^2+2(2x+y)+1}+2x+y} - \frac{1}{\sqrt{(x+y)^2+0.782\cdot 2(x+y)+0.782^2}+x+y}\right)}{\sqrt{\pi}} = \\
&\frac{2\sqrt{2}\sqrt{x}\left(\frac{2}{2(2x+y)+1} - \frac{1}{2(x+y)+0.782}\right)}{\sqrt{\pi}} = \\
&\frac{\left(2\sqrt{2}\sqrt{x}\right)\left(2(2(x+y)+0.782) - (2(2x+y)+1)\right)}{\sqrt{\pi}((2(x+y)+0.782)(2(2x+y)+1))} = \\
&\frac{\left(2\sqrt{2}\sqrt{x}\right)(2y+0.782\cdot 2-1)}{\sqrt{\pi}((2(x+y)+0.782)(2(2x+y)+1))} > 0 .
\end{aligned} \tag{222}$$

  We explain this chain of inequalities:

  - First inequality: We applied Lemma 22 two times.
  - Equalities factor out $\sqrt{2}\sqrt{x}$ and reformulate.
  - Second inequality part 1: we applied

$$0 < 2y \Longrightarrow (2x+y)^2 + 4x + 1 < (2x+y)^2 + 2(2x+y) + 1 = (2x+y+1)^2 . \tag{223}$$

– Second inequality part 2: we show that for $a = \frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}} - 13\right)$ following holds: $\frac{8x}{\pi} - \left(a^2 + 2a(x+y)\right) \geqslant 0$. We have $\frac{\partial}{\partial x}\frac{8x}{\pi} - \left(a^2 + 2a(x+y)\right) = \frac{8}{\pi} - 2a > 0$ and $\frac{\partial}{\partial y}\frac{8x}{\pi} - \left(a^2 + 2a(x+y)\right) = -2a < 0$. Therefore the minimum is at border for minimal x and maximal y:

$$\frac{8 \cdot 0.64}{\pi} - \left(\frac{2}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}} - 13\right)(0.64+0.01) + \left(\frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}} - 13\right)\right)^2\right) = 0\,. \tag{224}$$

Thus

$$\frac{8x}{\pi} \geqslant a^2 + 2a(x+y)\,. \tag{225}$$

for $a = \frac{1}{20}\left(\sqrt{\frac{2048+169\pi}{\pi}} - 13\right) > 0.782$.

– Equalities only solve square root and factor out the resulting terms $(2(2x+y)+1)$ and $(2(x+y)+0.782)$.

We know that $(2 - \mathrm{erfc}(x) > 0$ according to Lemma 21. For the sub-term we derived

$$2e^{\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2}\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) - e^{\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2}\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) > 0\,. \tag{226}$$

Consequently, both terms in the brackets of $\frac{\partial}{\partial \nu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ are larger than zero. Therefore $\frac{\partial}{\partial \nu}\tilde{\xi}(\mu,\omega,\nu,\tau,\lambda,\alpha)$ is larger than zero.

□

Lemma 41 (Mean at low variance)**.** *The mapping of the mean $\tilde{\mu}$ (Eq. (4))*

$$\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha) = \frac{1}{2}\lambda\left(-(\alpha+\mu\omega)\,\mathrm{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + \right. \tag{227}$$
$$\left. \alpha e^{\mu\omega+\frac{\nu\tau}{2}}\,\mathrm{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}e^{-\frac{\mu^2\omega^2}{2\nu\tau}} + 2\mu\omega\right)$$

in the domain $-0.1 \leqslant \mu \leqslant -0.1$, $-0.1 \leqslant \omega \leqslant -0.1$, and $0.02 \leqslant \nu\tau \leqslant 0.5$ is bounded by

$$|\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})| < 0.289324 \tag{228}$$

and

$$\lim_{\nu\to 0}|\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda_{01},\alpha_{01})| = \lambda\mu\omega. \tag{229}$$

We can consider $\tilde{\mu}$ with given $\mu\omega$ as a function in $x = \nu\tau$. We show the graph of this function at the maximal $\mu\omega = 0.01$ in the interval $x \in [0,1]$ in Figure A6.

Proof. Since $\tilde{\mu}$ is strictly monotonically increasing with $\mu\omega$

$$\tilde{\mu}(\mu,\omega,\nu,\tau,\lambda,\alpha) \leqslant \tag{230}$$
$$\tilde{\mu}(0.1,0.1,\nu,\tau,\lambda,\alpha) \leqslant$$
$$\frac{1}{2}\lambda\left(-(\alpha+0.01)\,\mathrm{erfc}\left(\frac{0.01}{\sqrt{2}\sqrt{\nu\tau}}\right) + \alpha e^{0.01+\frac{\nu\tau}{2}}\,\mathrm{erfc}\left(\frac{0.01+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) + \sqrt{\frac{2}{\pi}}\sqrt{\nu\tau}e^{-\frac{0.01^2}{2\nu\tau}} + 2\cdot 0.01\right) \leqslant$$
$$\frac{1}{2}\lambda_{01}\left(e^{\frac{0.05}{2}+0.01}\alpha_{01}\,\mathrm{erfc}\left(\frac{0.02+0.01}{\sqrt{2}\sqrt{0.02}}\right) - (\alpha_{01}+0.01)\,\mathrm{erfc}\left(\frac{0.01}{\sqrt{2}\sqrt{0.02}}\right) + e^{-\frac{0.01^2}{2\cdot 0.5}}\sqrt{0.5}\sqrt{\frac{2}{\pi}} + 0.01\cdot 2\right)$$
$$< 0.21857,$$

where we have used the monotonicity of the terms in $\nu\tau$.

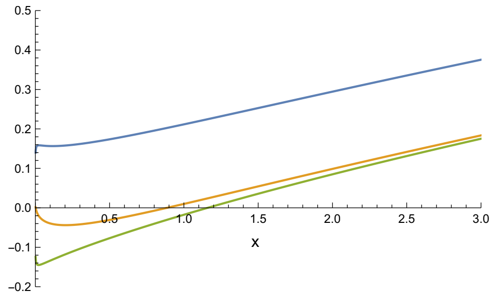

Figure A6: The graph of function $\tilde{\mu}$ for low variances $x = \nu\tau$ for $\mu\omega = 0.01$, where $x \in [0, 3]$, is displayed in yellow. Lower and upper bounds based on the Abramowitz bounds (Lemma 22) are displayed in green and blue, respectively.

Similarly, we can use the monotonicity of the terms in $\nu\tau$ to show that

$$\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda, \alpha) \geqslant \tilde{\mu}(0.1, -0.1, \nu, \tau, \lambda, \alpha) > -0.289324, \tag{231}$$

such that $|\tilde{\mu}| < 0.289324$ at low variances.

Furthermore, when $(\nu\tau) \to 0$, the terms with the arguments of the complementary error functions erfc and the exponential function go to infinity, therefore these three terms converge to zero. Hence, the remaining terms are only $2\mu\omega\frac{1}{2}\lambda$. $\square$

Lemma 42 (Bounds on derivatives of $\tilde{\mu}$ in Ω^-)**.** *The derivatives of the function $\tilde{\mu}(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}$ (Eq. (4)) with respect to μ, ω, ν, τ in the domain $\Omega^- = \{\mu, \omega, \nu, \tau \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.24, 0.8 \leqslant \tau \leqslant 1.25\}$ can be bounded as follows:*

$$\begin{aligned}
\left|\frac{\partial}{\partial \mu}\tilde{\mu}\right| &< 0.14 \qquad (232)\\
\left|\frac{\partial}{\partial \omega}\tilde{\mu}\right| &< 0.14\\
\left|\frac{\partial}{\partial \nu}\tilde{\mu}\right| &< 0.52\\
\left|\frac{\partial}{\partial \tau}\tilde{\mu}\right| &< 0.11.
\end{aligned}$$

Proof. The expression

$$\frac{\partial}{\partial \mu}\tilde{\mu} = J_{11} = \frac{1}{2}\lambda\omega e^{\frac{-(\mu\omega)^2}{2\nu\tau}}\left(2e^{\frac{(\mu\omega)^2}{2\nu\tau}} - e^{\frac{(\mu\omega)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right) + \alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{233}$$

contains the terms $e^{\frac{(\mu\omega)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega}{\sqrt{2}\sqrt{\nu\tau}}\right)$ and $e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)$ which are monotonically decreasing in their arguments (Lemma 23). We can therefore obtain their minima and maximal at the minimal and maximal arguments. Since the first term has a negative sign in the expression, both terms reach their maximal value at $\mu\omega = -0.01$, $\nu = 0.05$, and $\tau = 0.8$.

$$\left|\frac{\partial}{\partial \mu}\tilde{\mu}\right| \leqslant \frac{1}{2}|\lambda\omega|\left|\left(2 - e^{0.0353553^2}\operatorname{erfc}(0.0353553) + \alpha e^{0.106066^2}\operatorname{erfc}(0.106066)\right)\right| < 0.133 \tag{234}$$

Since, $\tilde{\mu}$ is symmetric in μ and ω, these bounds also hold for the derivate to ω.

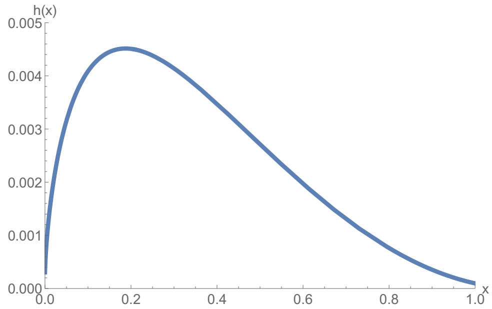

Figure A7: The graph of the function $h(x) = \tilde{\mu}^2(0.1, -0.1, x, 1, \lambda_{01}, \alpha_{01})$ is displayed. It has a local maximum at $x = \nu\tau \approx 0.187342$ and $h(x) \approx 0.00451457$ in the domain $x \in [0, 1]$.

We use the argumentation that the term with the error function is monotonically decreasing (Lemma 23) again for the expression

$$\frac{\partial}{\partial \nu}\tilde{\mu} = J_{12} = \tag{235}$$

$$= \frac{1}{4}\lambda\tau e^{-\frac{\mu^2\omega^2}{2\nu\tau}}\left(\alpha e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - (\alpha-1)\sqrt{\frac{2}{\pi\nu\tau}}\right) \leqslant$$

$$\left|\frac{1}{4}\lambda\tau\right| (|1.1072 - 2.68593|) < 0.52.$$

We have used that the term $1.1072 \leqslant \alpha_{01} e^{\frac{(\mu\omega+\nu\tau)^2}{2\nu\tau}} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) \leqslant 1.49042$ and the term $0.942286 \leqslant (\alpha - 1)\sqrt{\frac{2}{\pi\nu\tau}} \leqslant 2.68593$. Since $\tilde{\mu}$ is symmetric in ν and τ, we only have to chance outermost term $\left|\frac{1}{4}\lambda\tau\right|$ to $\left|\frac{1}{4}\lambda\nu\right|$ to obtain the estimate $\left|\frac{\partial}{\partial\tau}\tilde{\mu}\right| < 0.11$.

□

Lemma 43 (Tight bound on $\tilde{\mu}^2$ in Ω^-)**.** *The function $\tilde{\mu}^2(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01})$ (Eq. (4)) is bounded by*

$$\left|\tilde{\mu}^2\right| < 0.005 \tag{236}$$

$$\tag{237}$$

in the domain $\Omega^- = \{\mu, \omega, \nu, \tau \mid -0.1 \leqslant \mu \leqslant 0.1, -0.1 \leqslant \omega \leqslant 0.1, 0.05 \leqslant \nu \leqslant 0.24, 0.8 \leqslant \tau \leqslant 1.25\}$.

We visualize the function $\tilde{\mu}^2$ at its maximal $\mu\nu = -0.01$ and for $x = \nu\tau$ in the form $h(x) = \tilde{\mu}^2(0.1, -0.1, x, 1, \lambda_{01}, \alpha_{01})$ in Figure A7.

Proof. We use a similar strategy to the one we have used to show the bound on the singular value (Lemmata 10, 11, and 12), where we evaluted the function on a grid and used bounds on the derivatives together with the mean value theorem. Here we have

$$\left|\tilde{\mu}^2(\mu, \omega, \nu, \tau, \lambda_{01}, \alpha_{01}) - \tilde{\mu}^2(\mu + \Delta\mu, \omega + \Delta\omega, \nu + \Delta\nu, \tau + \Delta\tau, \lambda_{01}, \alpha_{01})\right| \leqslant \tag{238}$$

$$\left|\frac{\partial}{\partial\mu}\tilde{\mu}^2\right| |\Delta\mu| + \left|\frac{\partial}{\partial\omega}\tilde{\mu}^2\right| |\Delta\omega| + \left|\frac{\partial}{\partial\nu}\tilde{\mu}^2\right| |\Delta\nu| + \left|\frac{\partial}{\partial\tau}\tilde{\mu}^2\right| |\Delta\tau|.$$

We use Lemma 42 and Lemma 41, to obtain

$$\left|\frac{\partial}{\partial\mu}\tilde{\mu}^2\right| = 2\left|\tilde{\mu}\right| \left|\frac{\partial}{\partial\mu}\tilde{\mu}\right| < 2 \cdot 0.289324 \cdot 0.14 = 0.08101072 \tag{239}$$

$$\left|\frac{\partial}{\partial\omega}\tilde{\mu}^2\right| = 2\left|\tilde{\mu}\right| \left|\frac{\partial}{\partial\omega}\tilde{\mu}\right| < 2 \cdot 0.289324 \cdot 0.14 = 0.08101072$$

$$\left|\frac{\partial}{\partial \nu}\tilde{\mu}^2\right| = 2\,|\tilde{\mu}|\left|\frac{\partial}{\partial \nu}\tilde{\mu}\right| < 2 \cdot 0.289324 \cdot 0.52 = 0.30089696$$
$$\left|\frac{\partial}{\partial \tau}\tilde{\mu}^2\right| = 2\,|\tilde{\mu}|\left|\frac{\partial}{\partial \tau}\tilde{\mu}\right| < 2 \cdot 0.289324 \cdot 0.11 = 0.06365128$$

We evaluated the function $\tilde{\mu}^2$ in a grid G of Ω^- with $\Delta\mu = 0.001498041$, $\Delta\omega = 0.001498041$, $\Delta\nu = 0.0004033190$, and $\Delta\tau = 0.0019065994$ using a computer and obtained the maximal value $\max_G(\tilde{\mu})^2 = 0.00451457$, therefore the maximal value of $\tilde{\mu}^2$ is bounded by

$$\max_{(\mu,\omega,\nu,\tau)\in\Omega^-}(\tilde{\mu})^2 \leqslant \tag{240}$$
$$0.00451457 + 0.001498041 \cdot 0.08101072 + 0.001498041 \cdot 0.08101072 +$$
$$0.0004033190 \cdot 0.30089696 + 0.0019065994 \cdot 0.06365128 < 0.005. \tag{241}$$

Furthermore we used error propagation to estimate the numerical error on the function evaluation. Using the error propagation rules derived in Subsection A3.4.5, we found that the numerical error is smaller than 10^{-13} in the worst case. $\square$

Lemma 44 (Main subfunction)**.** *For* $1.2 \leqslant x \leqslant 20$ *and* $-0.1 \leqslant y \leqslant 0.1$,

the function

$$e^{\frac{(x+y)^2}{2x}}\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{242}$$

is smaller than zero, is strictly monotonically increasing in x*, and strictly monotonically decreasing in* y *for the minimal* $x = 12/10 = 1.2$*.*

Proof. We first consider the derivative of sub-function Eq. (101) with respect to x. The derivative of the function

$$e^{\frac{(x+y)^2}{2x}}\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{243}$$

with respect to x is

$$\frac{\sqrt{\pi}\left(e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right) + \sqrt{2}\sqrt{x}(3x-y)}{2\sqrt{\pi}x^2} = \tag{244}$$

$$\frac{\sqrt{\pi}\left(e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right) + \sqrt{2}\sqrt{x}(3x-y)}{2\sqrt{\pi}x^2} =$$

$$\frac{\sqrt{\pi}\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}} - \frac{2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right) + (3x-y)}{2\sqrt{2}\sqrt{\pi}x^2\sqrt{x}}.$$

We consider the numerator

$$\sqrt{\pi}\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}} - \frac{2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right) + (3x-y)\,. \tag{245}$$

For bounding this value, we use the approximation

$$e^{z^2}\,\mathrm{erfc}(z) \approx \frac{2.911}{\sqrt{\pi}(2.911-1)z + \sqrt{\pi z^2 + 2.911^2}}\,. \tag{246}$$

from Ren and MacKenzie [30]. We start with an error analysis of this approximation. According to Ren and MacKenzie [30] (Figure 1), the approximation error is positive in the range $[0.7, 3.2]$. This range contains all possible arguments of erfc that we consider. Numerically we maximized and minimized the approximation error of the whole expression

$$E(x,y) = \left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}} - \frac{2e^{\frac{(2x+y)^2}{2x}}(2x-y)(2x+y)\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right) - \tag{247}$$

$$\left(\frac{2.911(x-y)(x+y)}{\left(\sqrt{2}\sqrt{x}\right)\left(\frac{\sqrt{\pi}(2.911-1)(x+y)}{\sqrt{2}\sqrt{x}} + \sqrt{\pi\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2 + 2.911^2}\right)} - \right.$$

$$\left.\frac{2\cdot 2.911(2x-y)(2x+y)}{\left(\sqrt{2}\sqrt{x}\right)\left(\frac{\sqrt{\pi}(2.911-1)(2x+y)}{\sqrt{2}\sqrt{x}} + \sqrt{\pi\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2 + 2.911^2}\right)}\right).$$

We numerically determined $0.0113556 \leqslant E(x,y) \leqslant 0.0169551$ for $1.2 \leqslant x \leqslant 20$ and $-0.1 \leqslant y \leqslant 0.1$. We used different numerical optimization techniques like gradient based constraint BFGS algorithms and non-gradient-based Nelder-Mead methods with different start points. Therefore our approximation is smaller than the function that we approximate. We subtract an additional safety gap of 0.0131259 from our approximation to ensure that the inequality via the approximation holds true. With this safety gap the inequality would hold true even for negative x, where the approximation error becomes negative and the safety gap would compensate. Of course, the safety gap of 0.0131259 is not necessary for our analysis but may help or future investigations.

We have the sequences of inequalities using the approximation of Ren and MacKenzie [30]:

$$(3x-y) + \left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\,\mathrm{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}} - \frac{2e^{\frac{(2x+y)^2}{2x}}(2x-y)(2x+y)\,\mathrm{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)\sqrt{\pi} \geqslant \tag{248}$$

$$(3x-y) + \left(\frac{2.911(x-y)(x+y)}{\left(\sqrt{\pi\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2 + 2.911^2} + \frac{(2.911-1)\sqrt{\pi}(x+y)}{\sqrt{2}\sqrt{x}}\right)\left(\sqrt{2}\sqrt{x}\right)} - \right.$$

$$\left.\frac{2(2x-y)(2x+y)2.911}{\left(\sqrt{2}\sqrt{x}\right)\left(\sqrt{\pi\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2 + 2.911^2} + \frac{(2.911-1)\sqrt{\pi}(2x+y)}{\sqrt{2}\sqrt{x}}\right)}\right)\sqrt{\pi} - 0.0131259 =$$

$$(3x-y) + \left(\frac{\left(\sqrt{2}\sqrt{x}2.911\right)(x-y)(x+y)}{\left(\sqrt{\pi(x+y)^2 + 2\cdot 2.911^2 x} + (2.911-1)(x+y)\sqrt{\pi}\right)\left(\sqrt{2}\sqrt{x}\right)} - \right.$$

$$\left.\frac{2(2x-y)(2x+y)\left(\sqrt{2}\sqrt{x}2.911\right)}{\left(\sqrt{2}\sqrt{x}\right)\left(\sqrt{\pi(2x+y)^2 + 2\cdot 2.911^2 x} + (2.911-1)(2x+y)\sqrt{\pi}\right)}\right)\sqrt{\pi} - 0.0131259 =$$

$$
\begin{split}
&(3x-y) + 2.911 \left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{(x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} - \right.\\
&\left. \frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} \right) - 0.0131259 \ \geqslant \\
&(3x-y) + 2.911 \left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{\left(\frac{2.911^2}{\pi}\right)^2+(x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}+\frac{2\cdot 2.911^2 y}{\pi}}} - \right.\\
&\left. \frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} \right) - 0.0131259 \ = \\
&(3x-y) + 2.911 \left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{\left(x+y+\frac{2.911^2}{\pi}\right)^2}} - \right.\\
&\left. \frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} \right) - 0.0131259 \ = \\
&(3x-y) + 2.911 \left(\frac{(x-y)(x+y)}{2.911(x+y)+\frac{2.911^2}{\pi}} - \frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} \right) - 0.0131259 \ = \\
&(3x-y) + \frac{(x-y)(x+y)}{(x+y)+\frac{2.911}{\pi}} - \frac{2(2x-y)(2x+y)2.911}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}} - 0.0131259 \ = \qquad (3x-y)\\
&\left(-2(2x-y)2.911 \left((x+y)+\frac{2.911}{\pi}\right)(2x+y) \ + \right.\\
&\left((x+y)+\frac{2.911}{\pi}\right)(3x-y-0.0131259)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}\right) \ + \\
&\left. (x-y)(x+y)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}\right)\right)\\
&\left(\left((x+y)+\frac{2.911}{\pi}\right)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}\right)\right)^{-1} \ = \\
&\Big(((x-y)(x+y)+(3x-y-0.0131259)(x+y+0.9266))\left(\sqrt{(2x+y)^2+5.39467x}+3.822x+1.911y\right) \ - \\
&5.822(2x-y)(x+y+0.9266)(2x+y)\Big)\\
&\left(\left((x+y)+\frac{2.911}{\pi}\right)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{22.911^2 x}{\pi}}\right)\right)^{-1} \ > \ 0\,.
\end{split}
\tag{249}
$$

We explain this sequence of inequalities:

- First inequality: The approximation of Ren and MacKenzie [30] and then subtracting a safety gap (which would not be necessary for the current analysis).
- Equalities: The factor $\sqrt{2}\sqrt{x}$ is factored out and canceled.
- Second inequality: adds a positive term in the first root to obtain a binomial form. The term containing the root is positive and the root is in the denominator, therefore the whole term becomes smaller.

- Equalities: solve for the term and factor out.
- Bringing all terms to the denominator $\left((x+y)+\frac{2.911}{\pi}\right)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2 x}{\pi}}\right)$.
- Equalities: Multiplying out and expanding terms.
- Last inequality >0 is proofed in the following sequence of inequalities.

We look at the numerator of the last expression of Eq. (248), which we show to be positive in order to show >0 in Eq. (248). The numerator is

$$
\begin{aligned}
&((x-y)(x+y)+(3x-y-0.0131259)(x+y+0.9266))\left(\sqrt{(2x+y)^2+5.39467x}+3.822x+1.911y\right)- \qquad (250)\\
&5.822(2x-y)(x+y+0.9266)(2x+y)\ =\\
&-5.822(2x-y)(x+y+0.9266)(2x+y)+(3.822x+1.911y)((x-y)(x+y)+\\
&(3x-y-0.0131259)(x+y+0.9266))+((x-y)(x+y)+\\
&(3x-y-0.0131259)(x+y+0.9266))\sqrt{(2x+y)^2+5.39467x}\ =\\
&-8.0x^3+\left(4x^2+2xy+2.76667x-2y^2-0.939726y-0.0121625\right)\sqrt{(2x+y)^2+5.39467x}-\\
&8.0x^2y-11.0044x^2+2.0xy^2+1.69548xy-0.0464849x+2.0y^3+3.59885y^2-0.0232425y\ =\\
&-8.0x^3+\left(4x^2+2xy+2.76667x-2y^2-0.939726y-0.0121625\right)\sqrt{(2x+y)^2+5.39467x}-\\
&8.0x^2y-11.0044x^2+2.0xy^2+1.69548xy-0.0464849x+2.0y^3+3.59885y^2-0.0232425y\ .
\end{aligned}
$$

The factor in front of the root is positive. If the term, that does not contain the root, was positive, then the whole expression would be positive and we would have proofed that the numerator is positive. Therefore we consider the case that the term, that does not contain the root, is negative. The term that contains the root must be larger than the other term in absolute values.

$$
\begin{aligned}
&-\left(-8.0x^3-8.0x^2y-11.0044x^2+2.xy^2+1.69548xy-0.0464849x+2.y^3+3.59885y^2-0.0232425y\right)\ < \qquad (251)\\
&\left(4x^2+2xy+2.76667x-2y^2-0.939726y-0.0121625\right)\sqrt{(2x+y)^2+5.39467x}\ .
\end{aligned}
$$

Therefore the squares of the root term have to be larger than the square of the other term to show >0 in Eq. (248). Thus, we have the inequality:

$$
\begin{aligned}
&\left(-8.0x^3-8.0x^2y-11.0044x^2+2.xy^2+1.69548xy-0.0464849x+2.y^3+3.59885y^2-0.0232425y\right)^2\ < \qquad (252)\\
&\left(4x^2+2xy+2.76667x-2y^2-0.939726y-0.0121625\right)^2\left((2x+y)^2+5.39467x\right)\ .
\end{aligned}
$$

This is equivalent to

$$
\begin{aligned}
&0\ <\ \left(4x^2+2xy+2.76667x-2y^2-0.939726y-0.0121625\right)^2\left((2x+y)^2+5.39467x\right)- \qquad (253)\\
&\left(-8.0x^3-8.0x^2y-11.0044x^2+2.0xy^2+1.69548xy-0.0464849x+2.0y^3+3.59885y^2-0.0232425y\right)^2\ =\\
&-1.2227x^5+40.1006x^4y+27.7897x^4+41.0176x^3y^2+64.5799x^3y+39.4762x^3+10.9422x^2y^3-\\
&13.543x^2y^2-28.8455x^2y-0.364625x^2+0.611352xy^4+6.83183xy^3+5.46393xy^2+\\
&0.121746xy+0.000798008x-10.6365y^5-11.927y^4+0.190151y^3-0.000392287y^2\ .
\end{aligned}
$$

We obtain the inequalities:

$$
\begin{aligned}
&-1.2227x^5+40.1006x^4y+27.7897x^4+41.0176x^3y^2+64.5799x^3y+39.4762x^3+10.9422x^2y^3- \qquad (254)\\
&13.543x^2y^2-28.8455x^2y-0.364625x^2+0.611352xy^4+6.83183xy^3+5.46393xy^2+\\
&0.121746xy+0.000798008x-10.6365y^5-11.927y^4+0.190151y^3-0.000392287y^2\ =
\end{aligned}
$$

$$
\begin{aligned}
&-1.2227x^5 + 27.7897x^4 + 41.0176x^3y^2 + 39.4762x^3 - 13.543x^2y^2 - 0.364625x^2 + \\
&y\left(40.1006x^4 + 64.5799x^3 + 10.9422x^2y^2 - 28.8455x^2 + 6.83183xy^2 + 0.121746x - \right.\\
&\left.10.6365y^4 + 0.190151y^2\right) + 0.611352xy^4 + 5.46393xy^2 + 0.000798008x - 11.927y^4 - 0.000392287y^2 \; > \\
&-1.2227x^5 + 27.7897x^4 + 41.0176 \cdot (0.0)^2x^3 + 39.4762x^3 - 13.543 \cdot (0.1)^2x^2 - 0.364625x^2 - \\
&0.1 \cdot \left(40.1006x^4 + 64.5799x^3 + 10.9422 \cdot (0.1)^2x^2 - 28.8455x^2 + 6.83183 \cdot (0.1)^2x + 0.121746x + \right.\\
&\left.10.6365 \cdot (0.1)^4 + 0.190151 \cdot (0.1)^2\right) + \\
&0.611352 \cdot (0.0)^4x + 5.46393 \cdot (0.0)^2x + 0.000798008x - 11.927 \cdot (0.1)^4 - 0.000392287 \cdot (0.1)^2 \; = \\
&-1.2227x^5 + 23.7796x^4 + (20 + 13.0182)x^3 + 2.37355x^2 - 0.0182084x - 0.000194074 \; \geqslant \\
&-1.2227x^5 + 24.7796x^4 + 13.0182x^3 + 2.37355x^2 - 0.0182084x - 0.000194074 \; > \\
&13.0182x^3 + 2.37355x^2 - 0.0182084x - 0.000194074 \; > \; 0 \, .
\end{aligned}
$$

We used $24.7796 \cdot (20)^4 - 1.2227 \cdot (20)^5 = 52090.9 > 0$ and $x \leqslant 20$. We have proofed the last inequality > 0 of Eq. (248).

Consequently the derivative is always positive independent of y, thus

$$
e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{255}
$$

is strictly monotonically increasing in x.

The main subfunction is smaller than zero. Next we show that the sub-function Eq. (101) is smaller than zero. We consider the limit:

$$
\lim_{x \to \infty} e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) \; - \; 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \; = \; 0 \tag{256}
$$

The limit follows from Lemma 22. Since the function is monotonic increasing in x, it has to approach 0 from below. Thus,

$$
e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{257}
$$

is smaller than zero.

Behavior of the main subfunction with respect to y at minimal x. We now consider the derivative of sub-function Eq. (101) with respect to y. We proofed that sub-function Eq. (101) is strictly monotonically increasing independent of y. In the proof of Theorem 16, we need the minimum of sub-function Eq. (101). Therefore we are only interested in the derivative of sub-function Eq. (101) with respect to y for the minimum $x = 12/10 = 1.2$

Consequently, we insert the minimum $x = 12/10 = 1.2$ into the sub-function Eq. (101). The main terms become

$$
\frac{x+y}{\sqrt{2}\sqrt{x}} \; = \; \frac{y+1.2}{\sqrt{2}\sqrt{1.2}} \; = \; \frac{y}{\sqrt{2}\sqrt{1.2}} + \frac{\sqrt{1.2}}{\sqrt{2}} \; = \; \frac{5y+6}{2\sqrt{15}} \tag{258}
$$

and

$$
\frac{2x+y}{\sqrt{2}\sqrt{x}} \; = \; \frac{y+1.2 \cdot 2}{\sqrt{2}\sqrt{1.2}} \; = \; \frac{y}{\sqrt{2}\sqrt{1.2}} + \sqrt{1.2}\sqrt{2} \; = \; \frac{5y+12}{2\sqrt{15}} \, . \tag{259}
$$

Sub-function Eq. (101) becomes:

$$
e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{12}{10}}} + \frac{\sqrt{\frac{12}{10}}}{\sqrt{2}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{12}{10}}} + \frac{\sqrt{\frac{12}{10}}}{\sqrt{2}}\right) - 2e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{12}{10}}} + \sqrt{2}\sqrt{\frac{12}{10}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{12}{10}}} + \sqrt{2}\sqrt{\frac{12}{10}}\right) \, . \tag{260}
$$

The derivative of this function with respect to y is

$$\frac{\sqrt{15\pi}\left(e^{\frac{1}{60}(5y+6)^2}(5y+6)\operatorname{erfc}\left(\frac{5y+6}{2\sqrt{15}}\right)-2e^{\frac{1}{60}(5y+12)^2}(5y+12)\operatorname{erfc}\left(\frac{5y+12}{2\sqrt{15}}\right)\right)+30}{6\sqrt{15\pi}}\,. \tag{261}$$

We again will use the approximation of Ren and MacKenzie [30]

$$e^{z^2}\operatorname{erfc}(z) \;=\; \frac{2.911}{\sqrt{\pi}(2.911-1)z+\sqrt{\pi z^2+2.911^2}}\,. \tag{262}$$

Therefore we first perform an error analysis. We estimated the maximum and minimum of

$$\sqrt{15\pi}\left(\frac{2\cdot 2.911(5y+12)}{\frac{\sqrt{\pi}(2.911-1)(5y+12)}{2\sqrt{15}}+\sqrt{\pi\left(\frac{5y+12}{2\sqrt{15}}\right)^2+2.911^2}}-\frac{2.911(5y+6)}{\frac{\sqrt{\pi}(2.911-1)(5y+6)}{2\sqrt{15}}+\sqrt{\pi\left(\frac{5y+6}{2\sqrt{15}}\right)^2+2.911^2}}\right)+30+ \tag{263}$$

$$\sqrt{15\pi}\left(e^{\frac{1}{60}(5y+6)^2}(5y+6)\operatorname{erfc}\left(\frac{5y+6}{2\sqrt{15}}\right)-2e^{\frac{1}{60}(5y+12)^2}(5y+12)\operatorname{erfc}\left(\frac{5y+12}{2\sqrt{15}}\right)\right)+30\,.$$

We obtained for the maximal absolute error the value 0.163052. We added an approximation error of 0.2 to the approximation of the derivative. Since we want to show that the approximation upper bounds the true expression, the addition of the approximation error is required here. We get a sequence of inequalities:

$$\sqrt{15\pi}\left(e^{\frac{1}{60}(5y+6)^2}(5y+6)\operatorname{erfc}\left(\frac{5y+6}{2\sqrt{15}}\right)-2e^{\frac{1}{60}(5y+12)^2}(5y+12)\operatorname{erfc}\left(\frac{5y+12}{2\sqrt{15}}\right)\right)+30\;\leqslant \tag{264}$$

$$\sqrt{15\pi}\left(\frac{2.911(5y+6)}{\frac{\sqrt{\pi}(2.911-1)(5y+6)}{2\sqrt{15}}+\sqrt{\pi\left(\frac{5y+6}{2\sqrt{15}}\right)^2+2.911^2}}-\frac{2\cdot 2.911(5y+12)}{\frac{\sqrt{\pi}(2.911-1)(5y+12)}{2\sqrt{15}}+\sqrt{\pi\left(\frac{5y+12}{2\sqrt{15}}\right)^2+2.911^2}}\right)+$$

$$30+0.2\;=$$

$$\frac{(30\cdot 2.911)(5y+6)}{(2.911-1)(5y+6)+\sqrt{(5y+6)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}}-\frac{2(30\cdot 2.911)(5y+12)}{(2.911-1)(5y+12)+\sqrt{(5y+12)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}}+$$

$$30+0.2\;=$$

$$\left((0.2+30)\left((2.911-1)(5y+12)+\sqrt{(5y+12)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right)\right.$$

$$\left((2.911-1)(5y+6)+\sqrt{(5y+6)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right)-$$

$$2\cdot 30\cdot 2.911(5y+12)\left((2.911-1)(5y+6)+\sqrt{(5y+6)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right)+$$

$$\left.2.911\cdot 30(5y+6)\left((2.911-1)(5y+12)+\sqrt{(5y+12)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right)\right)$$

$$
\left(\left((2.911-1)(5y+6) + \sqrt{(5y+6)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) \right.
$$

$$
\left. \left((2.911-1)(5y+12) + \sqrt{(5y+12)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) \right)^{-1} \ < \ 0 \, .
$$

We explain this sequence of inequalities.

- First inequality: The approximation of Ren and MacKenzie [30] and then adding the error bound to ensure that the approximation is larger than the true value.
- First equality: The factor $2\sqrt{15}$ and $2\sqrt{\pi}$ are factored out and canceled.
- Second equality: Bringing all terms to the denominator

$$
\left((2.911-1)(5y+6) + \sqrt{(5y+6)^2 + \left(\frac{2\sqrt{15}2.911}{\sqrt{\pi}}\right)^2} \right) \tag{265}
$$

$$
\left((2.911-1)(5y+12) + \sqrt{(5y+12)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) .
$$

- Last inequality < 0 is proofed in the following sequence of inequalities.

We look at the numerator of the last term in Eq. (264). We have to proof that this numerator is smaller than zero in order to proof the last inequality of Eq. (264). The numerator is

$$
(0.2+30) \left((2.911-1)(5y+12) + \sqrt{(5y+12)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) \tag{266}
$$

$$
\left((2.911-1)(5y+6) + \sqrt{(5y+6)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) \ -
$$

$$
2\cdot 30\cdot 2.911(5y+12) \left((2.911-1)(5y+6) + \sqrt{(5y+6)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) +
$$

$$
2.911\cdot 30(5y+6) \left((2.911-1)(5y+12) + \sqrt{(5y+12)^2 + \left(\frac{2\sqrt{15}\, .2.911}{\sqrt{\pi}}\right)^2} \right) .
$$

We now compute upper bounds for this numerator:

$$
(0.2+30) \left((2.911-1)(5y+12) + \sqrt{(5y+12)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) \tag{267}
$$

$$
\left((2.911-1)(5y+6) + \sqrt{(5y+6)^2 + \left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2} \right) -
$$

$$
\begin{aligned}
&2 \cdot 30 \cdot 2.911(5y+12)\left((2.911-1)(5y+6)+\sqrt{(5y+6)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right)+\\
&2.911 \cdot 30(5y+6)\left((2.911-1)(5y+12)+\sqrt{(5y+12)^2+\left(\frac{2\sqrt{15}\cdot 2.911}{\sqrt{\pi}}\right)^2}\right) \;=\\
&-1414.99y^2-584.739\sqrt{(5y+6)^2+161.84}y+725.211\sqrt{(5y+12)^2+161.84}y-\\
&5093.97y-1403.37\sqrt{(5y+6)^2+161.84}+30.2\sqrt{(5y+6)^2+161.84}\sqrt{(5y+12)^2+161.84}+\\
&870.253\sqrt{(5y+12)^2+161.84}-4075.17 \;<\\
&-1414.99y^2-584.739\sqrt{(5y+6)^2+161.84}y+725.211\sqrt{(5y+12)^2+161.84}y-\\
&5093.97y-1403.37\sqrt{(6+5\cdot(-0.1))^2+161.84}+30.2\sqrt{(6+5\cdot 0.1)^2+161.84}\sqrt{(12+5\cdot 0.1)^2+161.84}+\\
&870.253\sqrt{(12+5\cdot 0.1)^2+161.84}-4075.17 \;=\\
&-1414.99y^2-584.739\sqrt{(5y+6)^2+161.84}y+725.211\sqrt{(5y+12)^2+161.84}y-5093.97y-309.691 \;<\\
&y\left(-584.739\sqrt{(5y+6)^2+161.84}+725.211\sqrt{(5y+12)^2+161.84}-5093.97\right)-309.691 \;<\\
&-0.1\left(725.211\sqrt{(12+5\cdot(-0.1))^2+161.84}-584.739\sqrt{(6+5\cdot 0.1)^2+161.84}-5093.97\right)-309.691 \;=\\
&-208.604\,.
\end{aligned}
$$

For the first inequality we choose y in the roots, so that positive terms maximally increase and negative terms maximally decrease. The second inequality just removed the y^2 term which is always negative, therefore increased the expression. For the last inequality, the term in brackets is negative for all settings of y. Therefore we make the brackets as negative as possible and make the whole term positive by multiplying with $y = -0.1$.

Consequently

$$
e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-2e^{\frac{(2x+y)^2}{2x}}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{268}
$$

is strictly monotonically decreasing in y for the minimal $x = 1.2$. □

Lemma 45 (Main subfunction below)**.** *For $0.007 \leqslant x \leqslant 0.875$ and $-0.01 \leqslant y \leqslant 0.01$, the function*

$$
e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-2e^{\frac{(2x+y)^2}{2x}}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{269}
$$

smaller than zero, is strictly monotonically increasing in x and strictly monotonically increasing in y for the minimal $x = 0.007 = 0.00875 \cdot 0.8$, $x = 0.56 = 0.7 \cdot 0.8$, $x = 0.128 = 0.16 \cdot 0.8$, and $x = 0.216 = 0.24 \cdot 0.9$ (lower bound of 0.9 on τ).

Proof. We first consider the derivative of sub-function Eq. (111) with respect to x. The derivative of the function

$$
e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-2e^{\frac{(2x+y)^2}{2x}}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{270}
$$

with respect to x is

$$
\frac{\sqrt{\pi}\left(e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-2e^{\frac{(2x+y)^2}{2x}}\left(4x^2-y^2\right)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right)+\sqrt{2}\sqrt{x}(3x-y)}{2\sqrt{\pi}x^2} = \tag{271}
$$

$$
\frac{\sqrt{\pi}\left(e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)-2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)\right)+\sqrt{2}\sqrt{x}(3x-y)}{2\sqrt{\pi}x^2} =
$$

$$\frac{\sqrt{\pi}\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}-\frac{2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)+(3x-y)}{\sqrt{2}2\sqrt{\pi}\sqrt{x}x^2}\,.$$

We consider the numerator

$$\sqrt{\pi}\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}-\frac{2e^{\frac{(2x+y)^2}{2x}}(2x+y)(2x-y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)+(3x-y)\,. \tag{272}$$

For bounding this value, we use the approximation

$$e^{z^2}\operatorname{erfc}(z)\;\approx\;\frac{2.911}{\sqrt{\pi}(2.911-1)z+\sqrt{\pi z^2+2.911^2}}\,. \tag{273}$$

from Ren and MacKenzie [30]. We start with an error analysis of this approximation. According to Ren and MacKenzie [30] (Figure 1), the approximation error is both positive and negative in the range $[0.175, 1.33]$. This range contains all possible arguments of erfc that we consider in this subsection. Numerically we maximized and minimized the approximation error of the whole expression

$$\begin{aligned}E(x,y)\;=\;&\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}-\frac{2e^{\frac{(2x+y)^2}{2x}}(2x-y)(2x+y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)-\\&\left(\frac{2.911(x-y)(x+y)}{\left(\sqrt{2}\sqrt{x}\right)\left(\frac{\sqrt{\pi}(2.911-1)(x+y)}{\sqrt{2}\sqrt{x}}+\sqrt{\pi\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2+2.911^2}\right)}-\right.\\&\left.\frac{2\cdot 2.911(2x-y)(2x+y)}{\left(\sqrt{2}\sqrt{x}\right)\left(\frac{\sqrt{\pi}(2.911-1)(2x+y)}{\sqrt{2}\sqrt{x}}+\sqrt{\pi\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2+2.911^2}\right)}\right)\,.\end{aligned} \tag{274}$$

We numerically determined $-0.000228141 \leqslant E(x,y) \leqslant 0.00495688$ for $0.08 \leqslant x \leqslant 0.875$ and $-0.01 \leqslant y \leqslant 0.01$. We used different numerical optimization techniques like gradient based constraint BFGS algorithms and non-gradient-based Nelder-Mead methods with different start points. Therefore our approximation is smaller than the function that we approximate.

We use an error gap of -0.0003 to countermand the error due to the approximation. We have the sequences of inequalities using the approximation of Ren and MacKenzie [30]:

$$\begin{aligned}&(3x-y)+\left(\frac{e^{\frac{(x+y)^2}{2x}}(x-y)(x+y)\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}-\frac{2e^{\frac{(2x+y)^2}{2x}}(2x-y)(2x+y)\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)}{\sqrt{2}\sqrt{x}}\right)\sqrt{\pi}\;\geqslant\\&(3x-y)+\left(\frac{2.911(x-y)(x+y)}{\left(\sqrt{\pi\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)^2+2.911^2}+\frac{(2.911-1)\sqrt{\pi}(x+y)}{\sqrt{2}\sqrt{x}}\right)\left(\sqrt{2}\sqrt{x}\right)}-\right.\end{aligned} \tag{275}$$

$$\left.\frac{2(2x-y)(2x+y)2.911}{\left(\sqrt{2}\sqrt{x}\right)\left(\sqrt{\pi\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right)^2+2.911^2}+\frac{(2.911-1)\sqrt{\pi}(2x+y)}{\sqrt{2}\sqrt{x}}\right)}\right)\sqrt{\pi}-0.0003 \ =$$

$$(3x-y)+\left(\frac{\left(\sqrt{2}\sqrt{x}2.911\right)(x-y)(x+y)}{\left(\sqrt{\pi(x+y)^2+2\cdot 2.911^2x}+(2.911-1)(x+y)\sqrt{\pi}\right)\left(\sqrt{2}\sqrt{x}\right)}-\right.$$

$$\left.\frac{2(2x-y)(2x+y)\left(\sqrt{2}\sqrt{x}2.911\right)}{\left(\sqrt{2}\sqrt{x}\right)\left(\sqrt{\pi(2x+y)^2+2\cdot 2.911^2x}+(2.911-1)(2x+y)\sqrt{\pi}\right)}\right)\sqrt{\pi}-0.0003 \ =$$

$$(3x-y)+2.911\left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{(x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}-\right.$$

$$\left.\frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}\right)-0.0003 \ \geqslant$$

$$(3x-y)+2.911\left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{\left(\frac{2.911^2}{\pi}\right)^2+(x+y)^2+\frac{2\cdot 2.911^2x}{\pi}+\frac{2\cdot 2.911^2y}{\pi}}}-\right.$$

$$\left.\frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}\right)-0.0003 \ =$$

$$(3x-y)+2.911\left(\frac{(x-y)(x+y)}{(2.911-1)(x+y)+\sqrt{\left(x+y+\frac{2.911^2}{\pi}\right)^2}}-\right.$$

$$\left.\frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}\right)-0.0003 \ =$$

$$(3x-y)+2.911\left(\frac{(x-y)(x+y)}{2.911(x+y)+\frac{2.911^2}{\pi}}-\frac{2(2x-y)(2x+y)}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}\right)-0.0003 \ =$$

$$(3x-y)+\frac{(x-y)(x+y)}{(x+y)+\frac{2.911}{\pi}}-\frac{2(2x-y)(2x+y)2.911}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}-0.0003 \ =$$

$$(3x-y)+\frac{(x-y)(x+y)}{(x+y)+\frac{2.911}{\pi}}-\frac{2(2x-y)(2x+y)2.911}{(2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}}-0.0003 \ =$$

$$\left(-2(2x-y)2.911\left((x+y)+\frac{2.911}{\pi}\right)(2x+y) \ +\right.$$

$$\left((x+y)+\frac{2.911}{\pi}\right)(3x-y-0.0003)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}\right) \ +$$

$$\left.(x-y)(x+y)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}\right)\right)$$

$$\left(\left((x+y)+\frac{2.911}{\pi}\right)\left((2.911-1)(2x+y)+\sqrt{(2x+y)^2+\frac{2\cdot 2.911^2x}{\pi}}\right)\right)^{-1} \ =$$

$$\Big(-8x^3 - 8x^2y + 4x^2\sqrt{(2x+y)^2 + 5.39467x} - 10.9554x^2 + 2xy^2 - 2y^2\sqrt{(2x+y)^2 + 5.39467x} +$$
$$1.76901xy + 2xy\sqrt{(2x+y)^2 + 5.39467x} + 2.7795x\sqrt{(2x+y)^2 + 5.39467x} -$$
$$0.9269y\sqrt{(2x+y)^2 + 5.39467x} - 0.00027798\sqrt{(2x+y)^2 + 5.39467x} - 0.00106244x +$$
$$2y^3 + 3.62336y^2 - 0.00053122y\Big)$$
$$\left(\left((x+y) + \frac{2.911}{\pi}\right)\left((2.911-1)(2x+y) + \sqrt{(2x+y)^2 + \frac{2 \cdot 2.911^2 x}{\pi}}\right)\right)^{-1} =$$
$$\Big(-8x^3 + \big(4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798\big)\sqrt{(2x+y)^2 + 5.39467x} -$$
$$8x^2y - 10.9554x^2 + 2xy^2 + 1.76901xy - 0.00106244x + 2y^3 + 3.62336y^2 - 0.00053122y\Big)$$
$$\left(\left((x+y) + \frac{2.911}{\pi}\right)\left((2.911-1)(2x+y) + \sqrt{(2x+y)^2 + \frac{2 \cdot 2.911^2 x}{\pi}}\right)\right)^{-1} > 0 .$$

We explain this sequence of inequalities:

- First inequality: The approximation of Ren and MacKenzie [30] and then subtracting an error gap of 0.0003.
- Equalities: The factor $\sqrt{2}\sqrt{x}$ is factored out and canceled.
- Second inequality: adds a positive term in the first root to obtain a binomial form. The term containing the root is positive and the root is in the denominator, therefore the whole term becomes smaller.
- Equalities: solve for the term and factor out.
- Bringing all terms to the denominator $\left((x+y) + \frac{2.911}{\pi}\right)\left((2.911-1)(2x+y) + \sqrt{(2x+y)^2 + \frac{2 \cdot 2.911^2 x}{\pi}}\right)$.
- Equalities: Multiplying out and expanding terms.
- Last inequality > 0 is proofed in the following sequence of inequalities.

We look at the numerator of the last expression of Eq. (275), which we show to be positive in order to show > 0 in Eq. (275). The numerator is

$$-8x^3 + \big(4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798\big)\sqrt{(2x+y)^2 + 5.39467x} - \tag{276}$$
$$8x^2y - 10.9554x^2 + 2xy^2 + 1.76901xy - 0.00106244x + 2y^3 + 3.62336y^2 - 0.00053122y .$$

The factor $4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798$ in front of the root is positive:

$$4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798 > \tag{277}$$
$$-2y^2 + 0.007 \cdot 2y - 0.9269y + 4 \cdot 0.007^2 + 2.7795 \cdot 0.007 - 0.00027798 =$$
$$-2y^2 - 0.9129y + 2.77942 = -2(y + 1.42897)(y - 0.972523) > 0 .$$

If the term that does not contain the root would be positive, then everything is positive and we have proofed the the numerator is positive. Therefore we consider the case that the term that does not contain the root is negative. The term that contains the root must be larger than the other term in absolute values.

$$-\big(-8x^3 - 8x^2y - 10.9554x^2 + 2xy^2 + 1.76901xy - 0.00106244x + 2y^3 + 3.62336y^2 - 0.00053122y\big) < \tag{278}$$
$$\big(4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798\big)\sqrt{(2x+y)^2 + 5.39467x} .$$

Therefore the squares of the root term have to be larger than the square of the other term to show > 0 in Eq. (275). Thus, we have the inequality:

$$\big(-8x^3 - 8x^2y - 10.9554x^2 + 2xy^2 + 1.76901xy - 0.00106244x + 2y^3 + 3.62336y^2 - 0.00053122y\big)^2 < \tag{279}$$

$$\left(4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798\right)^2 \left((2x + y)^2 + 5.39467x\right) .$$

This is equivalent to

$$0 < \left(4x^2 + 2xy + 2.7795x - 2y^2 - 0.9269y - 0.00027798\right)^2 \left((2x + y)^2 + 5.39467x\right) - \tag{280}$$
$$\left(-8x^3 - 8x^2y - 10.9554x^2 + 2xy^2 + 1.76901xy - 0.00106244x + 2y^3 + 3.62336y^2 - 0.00053122y\right)^2 =$$
$$x \cdot 4.168614250 \cdot 10^{-7} - y^2 2.049216091 \cdot 10^{-7} - 0.0279456x^5 +$$
$$43.0875x^4y + 30.8113x^4 + 43.1084x^3y^2 + 68.989x^3y + 41.6357x^3 + 10.7928x^2y^3 - 13.1726x^2y^2 -$$
$$27.8148x^2y - 0.00833715x^2 + 0.0139728xy^4 + 5.47537xy^3 +$$
$$4.65089xy^2 + 0.00277916xy - 10.7858y^5 - 12.2664y^4 + 0.00436492y^3 .$$

We obtain the inequalities:

$$x \cdot 4.168614250 \cdot 10^{-7} - y^2 2.049216091 \cdot 10^{-7} - 0.0279456x^5 + \tag{281}$$
$$43.0875x^4y + 30.8113x^4 + 43.1084x^3y^2 + 68.989x^3y + 41.6357x^3 + 10.7928x^2y^3 -$$
$$13.1726x^2y^2 - 27.8148x^2y - 0.00833715x^2 +$$
$$0.0139728xy^4 + 5.47537xy^3 + 4.65089xy^2 + 0.00277916xy - 10.7858y^5 - 12.2664y^4 + 0.00436492y^3 >$$
$$x \cdot 4.168614250 \cdot 10^{-7} - (0.01)^2 2.049216091 \cdot 10^{-7} - 0.0279456x^5 +$$
$$0.0 \cdot 43.0875x^4 + 30.8113x^4 + 43.1084(0.0)^2x^3 + 0.0 \cdot 68.989x^3 + 41.6357x^3 +$$
$$10.7928(0.0)^3x^2 - 13.1726(0.01)^2x^2 - 27.8148(0.01)x^2 - 0.00833715x^2 +$$
$$0.0139728(0.0)^4x + 5.47537(0.0)^3x + 4.65089(0.0)^2x +$$
$$0.0 \cdot 0.00277916x - 10.7858(0.01)^5 - 12.2664(0.01)^4 + 0.00436492(0.0)^3 =$$
$$x \cdot 4.168614250 \cdot 10^{-7} - 1.237626189 \cdot 10^{-7} - 0.0279456x^5 + 30.8113x^4 + 41.6357x^3 - 0.287802x^2 >$$
$$-\left(\frac{x}{0.007}\right)^3 1.237626189 \cdot 10^{-7} + 30.8113x^4 - (0.875) \cdot 0.0279456x^4 + 41.6357x^3 - \frac{(0.287802x)x^2}{0.007} =$$
$$30.7869x^4 + 0.160295x^3 > 0 .$$

We used $x \geqslant 0.007$ and $x \leqslant 0.875$ (reducing the negative x^4-term to a x^3-term). We have proofed the last inequality > 0 of Eq. (275).

Consequently the derivative is always positive independent of y, thus

$$e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x + y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x + y}{\sqrt{2}\sqrt{x}}\right) \tag{282}$$

is strictly monotonically increasing in x.

Next we show that the sub-function Eq. (111) is smaller than zero. We consider the limit:

$$\lim_{x\to\infty} e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x + y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x + y}{\sqrt{2}\sqrt{x}}\right) = 0 \tag{283}$$

The limit follows from Lemma 22. Since the function is monotonic increasing in x, it has to approach 0 from below. Thus,

$$e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x + y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x + y}{\sqrt{2}\sqrt{x}}\right) \tag{284}$$

is smaller than zero.

We now consider the derivative of sub-function Eq. (111) with respect to y. We proofed that sub-function Eq. (111) is strictly monotonically increasing independent of y. In the proof of Theorem 3, we need the minimum of sub-function Eq. (111). First, we are interested in the derivative of sub-function Eq. (111) with respect to y for the minimum $x = 0.007 = 7/1000$.

Consequently, we insert the minimum $x = 0.007 = 7/1000$ into the sub-function Eq. (111):

$$e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{7}{1000}}}+\frac{\sqrt{\frac{7}{1000}}}{\sqrt{2}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{7}{1000}}}+\frac{\sqrt{\frac{7}{1000}}}{\sqrt{2}}\right) - \tag{285}$$
$$2e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{7}{1000}}}+\sqrt{2}\sqrt{\frac{7}{1000}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{7}{1000}}}+\sqrt{2}\sqrt{\frac{7}{1000}}\right) =$$
$$e^{\frac{500y^2}{7}+y+\frac{7}{2000}} \operatorname{erfc}\left(\frac{1000y+7}{20\sqrt{35}}\right) - 2e^{\frac{(500y+7)^2}{3500}} \operatorname{erfc}\left(\frac{500y+7}{10\sqrt{35}}\right) .$$

The derivative of this function with respect to y is

$$\left(\frac{1000y}{7}+1\right) e^{\frac{500y^2}{7}+y+\frac{7}{2000}} \operatorname{erfc}\left(\frac{1000y+7}{20\sqrt{35}}\right) - \tag{286}$$
$$\frac{1}{7}4e^{\frac{(500y+7)^2}{3500}}(500y+7) \operatorname{erfc}\left(\frac{500y+7}{10\sqrt{35}}\right) + 20\sqrt{\frac{5}{7\pi}} >$$
$$\left(1+\frac{1000\cdot(-0.01)}{7}\right) e^{-0.01+\frac{7}{2000}+\frac{500\cdot(-0.01)^2}{7}} \operatorname{erfc}\left(\frac{7+1000+(-0.01)}{20\sqrt{35}}\right) -$$
$$\frac{1}{7}4e^{\frac{(7+500\cdot 0.01)^2}{3500}}(7+500\cdot 0.01) \operatorname{erfc}\left(\frac{7+500\cdot 0.01}{10\sqrt{35}}\right) + 20\sqrt{\frac{5}{7\pi}} > 3.56 .$$

For the first inequality, we use Lemma 24. Lemma 24 says that the function $xe^{x^2}\operatorname{erfc}(x)$ has the sign of x and is monotonically increasing to $\frac{1}{\sqrt{\pi}}$. Consequently, we inserted the maximal $y = 0.01$ to make the negative term more negative and the minimal $y = -0.01$ to make the positive term less positive.

Consequently

$$e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{287}$$

is strictly monotonically increasing in y for the minimal $x = 0.007$.

Next, we consider $x = 0.7 \cdot 0.8 = 0.56$, which is the maximal $\nu = 0.7$ and minimal $\tau = 0.8$. We insert the minimum $x = 0.56 = 56/100$ into the sub-function Eq. (111):

$$e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{56}{100}}}+\frac{\sqrt{\frac{56}{100}}}{\sqrt{2}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{56}{100}}}+\frac{\sqrt{\frac{56}{100}}}{\sqrt{2}}\right) - \tag{288}$$
$$2e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{56}{100}}}+\sqrt{2}\sqrt{\frac{56}{100}}\right)^2} \operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{56}{100}}}+\sqrt{2}\sqrt{\frac{56}{100}}\right) .$$

The derivative with respect to y is:

$$\frac{5e^{\left(\frac{5y}{2\sqrt{7}}+\frac{\sqrt{7}}{5}\right)^2}\left(\frac{5y}{2\sqrt{7}}+\frac{\sqrt{7}}{5}\right)\operatorname{erfc}\left(\frac{5y}{2\sqrt{7}}+\frac{\sqrt{7}}{5}\right)}{\sqrt{7}} - \tag{289}$$
$$\frac{10e^{\left(\frac{5y}{2\sqrt{7}}+\frac{2\sqrt{7}}{5}\right)^2}\left(\frac{5y}{2\sqrt{7}}+\frac{2\sqrt{7}}{5}\right)\operatorname{erfc}\left(\frac{5y}{2\sqrt{7}}+\frac{2\sqrt{7}}{5}\right)}{\sqrt{7}} + \frac{5}{\sqrt{7\pi}} >$$
$$\frac{5e^{\left(\frac{\sqrt{7}}{5}-\frac{0.01\cdot 5}{2\sqrt{7}}\right)^2}\left(\frac{\sqrt{7}}{5}-\frac{0.01\cdot 5}{2\sqrt{7}}\right)\operatorname{erfc}\left(\frac{\sqrt{7}}{5}-\frac{0.01\cdot 5}{2\sqrt{7}}\right)}{\sqrt{7}} -$$

$$\frac{10e^{\left(\frac{2\sqrt{7}}{5}+\frac{0.01\cdot 5}{2\sqrt{7}}\right)^2}\left(\frac{2\sqrt{7}}{5}+\frac{0.01\cdot 5}{2\sqrt{7}}\right)\operatorname{erfc}\left(\frac{2\sqrt{7}}{5}+\frac{0.01\cdot 5}{2\sqrt{7}}\right)}{\sqrt{7}}+\frac{5}{\sqrt{7\pi}} > 0.00746\,.$$

For the first inequality we applied Lemma 24 which states that the function $xe^{x^2}\operatorname{erfc}(x)$ is monotonically increasing. Consequently, we inserted the maximal $y = 0.01$ to make the negative term more negative and the minimal $y = -0.01$ to make the positive term less positive.

Consequently

$$e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{290}$$

is strictly monotonically increasing in y for $x = 0.56$.

Next, we consider $x = 0.16 \cdot 0.8 = 0.128$, which is the minimal $\tau = 0.8$. We insert the minimum $x = 0.128 = 128/1000$ into the sub-function Eq. (111):

$$\begin{aligned}
&e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{128}{1000}}}+\frac{\sqrt{\frac{128}{1000}}}{\sqrt{2}}\right)^2}\operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{128}{1000}}}+\frac{\sqrt{\frac{128}{1000}}}{\sqrt{2}}\right) - \\
&2e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{128}{1000}}}+\sqrt{2}\sqrt{\frac{128}{1000}}\right)^2}\operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{128}{1000}}}+\sqrt{2}\sqrt{\frac{128}{1000}}\right) = \\
&e^{\frac{125y^2}{32}+y+\frac{8}{125}}\operatorname{erfc}\left(\frac{125y+16}{20\sqrt{10}}\right) - 2e^{\frac{(125y+32)^2}{4000}}\operatorname{erfc}\left(\frac{125y+32}{20\sqrt{10}}\right)\,.
\end{aligned} \tag{291}$$

The derivative with respect to y is:

$$\begin{aligned}
&\frac{1}{16}\left(e^{\frac{125y^2}{32}+y+\frac{8}{125}}(125y+16)\operatorname{erfc}\left(\frac{125y+16}{20\sqrt{10}}\right) - \right.\\
&\left.2e^{\frac{(125y+32)^2}{4000}}(125y+32)\operatorname{erfc}\left(\frac{125y+32}{20\sqrt{10}}\right)+20\sqrt{\frac{10}{\pi}}\right) > \\
&\frac{1}{16}\left((16+125(-0.01))e^{-0.01+\frac{8}{125}+\frac{125(-0.01)^2}{32}}\operatorname{erfc}\left(\frac{16+125(-0.01)}{20\sqrt{10}}\right) - \right.\\
&\left.2e^{\frac{(32+1250.01)^2}{4000}}(32+1250.01)\operatorname{erfc}\left(\frac{32+1250.01}{20\sqrt{10}}\right)+20\sqrt{\frac{10}{\pi}}\right) > 0.4468\,.
\end{aligned} \tag{292}$$

For the first inequality we applied Lemma 24 which states that the function $xe^{x^2}\operatorname{erfc}(x)$ is monotonically increasing. Consequently, we inserted the maximal $y = 0.01$ to make the negative term more negative and the minimal $y = -0.01$ to make the positive term less positive.

Consequently

$$e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}}\operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{293}$$

is strictly monotonically increasing in y for $x = 0.128$.

Next, we consider $x = 0.24 \cdot 0.9 = 0.216$, which is the minimal $\tau = 0.9$ (here we consider 0.9 as lower bound for τ). We insert the minimum $x = 0.216 = 216/1000$ into the sub-function Eq. (111):

$$\begin{aligned}
&e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{216}{1000}}}+\frac{\sqrt{\frac{216}{1000}}}{\sqrt{2}}\right)^2}\operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{216}{1000}}}+\frac{\sqrt{\frac{216}{1000}}}{\sqrt{2}}\right) - \\
&2e^{\left(\frac{y}{\sqrt{2}\sqrt{\frac{216}{1000}}}+\sqrt{2}\sqrt{\frac{216}{1000}}\right)^2}\operatorname{erfc}\left(\frac{y}{\sqrt{2}\sqrt{\frac{216}{1000}}}+\sqrt{2}\sqrt{\frac{216}{1000}}\right) =
\end{aligned} \tag{294}$$

$$e^{\frac{(125y+27)^2}{6750}} \operatorname{erfc}\left(\frac{125y+27}{15\sqrt{30}}\right) - 2e^{\frac{(125y+54)^2}{6750}} \operatorname{erfc}\left(\frac{125y+54}{15\sqrt{30}}\right)$$

The derivative with respect to y is:

$$\frac{1}{27}\left(e^{\frac{(125y+27)^2}{6750}}(125y+27) \operatorname{erfc}\left(\frac{125y+27}{15\sqrt{30}}\right) - \right. \tag{295}$$
$$\left. 2e^{\frac{(125y+54)^2}{6750}}(125y+54) \operatorname{erfc}\left(\frac{125y+54}{15\sqrt{30}}\right) + 15\sqrt{\frac{30}{\pi}}\right) >$$
$$\frac{1}{27}\left((27+125(-0.01))e^{\frac{(27+125(-0.01))^2}{6750}} \operatorname{erfc}\left(\frac{27+125(-0.01)}{15\sqrt{30}}\right) - \right.$$
$$\left. 2e^{\frac{(54+1250.01)^2}{6750}}(54+1250.01) \operatorname{erfc}\left(\frac{54+1250.01}{15\sqrt{30}}\right) + 15\sqrt{\frac{30}{\pi}}\right)) > 0.211288 .$$

For the first inequality we applied Lemma 24 which states that the function $xe^{x^2}\operatorname{erfc}(x)$ is monotonically increasing. Consequently, we inserted the maximal $y = 0.01$ to make the negative term more negative and the minimal $y = -0.01$ to make the positive term less positive.

Consequently

$$e^{\frac{(x+y)^2}{2x}} \operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right) - 2e^{\frac{(2x+y)^2}{2x}} \operatorname{erfc}\left(\frac{2x+y}{\sqrt{2}\sqrt{x}}\right) \tag{296}$$

is strictly monotonically increasing in y for $x = 0.216$. □

Lemma 46 (Monotone Derivative)**.** *For $\lambda = \lambda_{01}$, $\alpha = \alpha_{01}$ and the domain $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$, $0.00875 \leqslant \nu \leqslant 0.7$, and $0.8 \leqslant \tau \leqslant 1.25$. We are interested of the derivative of*

$$\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2e^{\left(\frac{\mu\omega+2\cdot\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\cdot\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) . \tag{297}$$

The derivative of the equation above with respect to

- *ν is larger than zero;*
- *τ is smaller than zero for maximal $\nu = 0.7$, $\nu = 0.16$, and $\nu = 0.24$ (with $0.9 \leqslant \tau$);*
- *$y = \mu\omega$ is larger than zero for $\nu\tau = 0.00875 \cdot 0.8 = 0.007$, $\nu\tau = 0.7 \cdot 0.8 = 0.56$, $\nu\tau = 0.16 \cdot 0.8 = 0.128$, and $\nu\tau = 0.24 \cdot 0.9 = 0.216$.*

Proof. We consider the domain: $-0.1 \leqslant \mu \leqslant 0.1$, $-0.1 \leqslant \omega \leqslant 0.1$, $0.00875 \leqslant \nu \leqslant 0.7$, and $0.8 \leqslant \tau \leqslant 1.25$.

We use Lemma 17 to determine the derivatives. Consequently, the derivative of

$$\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{298}$$

with respect to ν is larger than zero, which follows directly from Lemma 17 using the chain rule.

Consequently, the derivative of

$$\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right) - 2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2} \operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{299}$$

with respect to $y = \mu\omega$ is larger than zero for $\nu\tau = 0.00875 \cdot 0.8 = 0.007$, $\nu\tau = 0.7 \cdot 0.8 = 0.56$, $\nu\tau = 0.16 \cdot 0.8 = 0.128$, and $\nu\tau = 0.24 \cdot 0.9 = 0.216$, which also follows directly from Lemma 17.

We now consider the derivative with respect to τ, which is not trivial since τ is a factor of the whole expression. The sub-expression should be maximized as it appears with negative sign in the mapping for ν.

First, we consider the function for the largest $\nu = 0.7$ and the largest $y = \mu\omega = 0.01$ for determining the derivative with respect to τ.

The expression becomes

$$\tau\left(e^{\left(\frac{\frac{7\tau}{10}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{7\tau}{10}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{7\tau}{10}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{7\tau}{10}}}\right)-2e^{\left(\frac{\frac{2\cdot 7\tau}{10}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{7\tau}{10}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{2\cdot 7\tau}{10}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{7\tau}{10}}}\right)\right) . \tag{300}$$

The derivative with respect to τ is

$$\begin{aligned}&\left(\sqrt{\pi}\left(e^{\frac{(70\tau+1)^2}{14000\tau}}(700\tau(7\tau+20)-1)\operatorname{erfc}\left(\frac{70\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)-\right.\right.\\&\left.\left.2e^{\frac{(140\tau+1)^2}{14000\tau}}(2800\tau(7\tau+5)-1)\operatorname{erfc}\left(\frac{140\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)\right)+20\sqrt{35}(210\tau-1)\sqrt{\tau}\right)\\&\left(14000\sqrt{\pi}\tau\right)^{-1} .\end{aligned} \tag{301}$$

We are considering only the numerator and use again the approximation of Ren and MacKenzie [30]. The error analysis on the whole numerator gives an approximation error $97 < E < 186$. Therefore we add 200 to the numerator when we use the approximation Ren and MacKenzie [30]. We obtain the inequalities:

$$\begin{aligned}&\sqrt{\pi}\left(e^{\frac{(70\tau+1)^2}{14000\tau}}(700\tau(7\tau+20)-1)\operatorname{erfc}\left(\frac{70\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)-\right.\\&\left.2e^{\frac{(140\tau+1)^2}{14000\tau}}(2800\tau(7\tau+5)-1)\operatorname{erfc}\left(\frac{140\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)\right)+20\sqrt{35}(210\tau-1)\sqrt{\tau}\ \leqslant\\&\sqrt{\pi}\left(\frac{2.911(700\tau(7\tau+20)-1)}{\frac{\sqrt{\pi}(2.911-1)(70\tau+1)}{20\sqrt{35}\sqrt{\tau}}+\sqrt{\pi\left(\frac{70\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)^2+2.911^2}}-\right.\\&\left.\frac{2\cdot 2.911(2800\tau(7\tau+5)-1)}{\frac{\sqrt{\pi}(2.911-1)(140\tau+1)}{20\sqrt{35}\sqrt{\tau}}+\sqrt{\pi\left(\frac{140\tau+1}{20\sqrt{35}\sqrt{\tau}}\right)^2+2.911^2}}\right)\\&+20\sqrt{35}(210\tau-1)\sqrt{\tau}+200\ =\\&\sqrt{\pi}\left(\frac{(700\tau(7\tau+20)-1)\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\cdot 2.911\sqrt{35}\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}}-\right.\\&\left.\frac{2(2800\tau(7\tau+5)-1)\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}}\right)+\\&\left(20\sqrt{35}(210\tau-1)\sqrt{\tau}+200\right)\ =\\&\left(\left(20\sqrt{35}(210\tau-1)\sqrt{\tau}+200\right)\left(\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}\right)\right.\\&\left(\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}\right)+\\&2.911\cdot 20\sqrt{35}\sqrt{\pi}(700\tau(7\tau+20)-1)\sqrt{\tau}\\&\left(\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}\right)-\end{aligned} \tag{302}$$

$$\sqrt{\pi}2 \cdot 20 \cdot \sqrt{35} \cdot 2.911(2800\tau(7\tau+5)-1)$$
$$\sqrt{\tau}\left(\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}\right)\Bigg)$$
$$\Bigg(\Bigg(\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\sqrt{35}\cdot 2.911\cdot\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}\Bigg)$$
$$\left(\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\sqrt{35}\cdot 2.911\cdot\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}\right)\Bigg)^{-1}.$$

After applying the approximation of Ren and MacKenzie [30] and adding 200, we first factored out $20\sqrt{35}\sqrt{\tau}$. Then we brought all terms to the same denominator.

We now consider the numerator:

$$\left(20\sqrt{35}(210\tau-1)\sqrt{\tau}+200\right)\left(\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}\right) \tag{303}$$
$$\left(\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}\right)+$$
$$2.911\cdot 20\sqrt{35}\sqrt{\pi}(700\tau(7\tau+20)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(140\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(140\tau+1)^2}\right)-$$
$$\sqrt{\pi}2\cdot 20\cdot\sqrt{35}\cdot 2.911(2800\tau(7\tau+5)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(70\tau+1)+\sqrt{\left(20\cdot\sqrt{35}\cdot 2.911\sqrt{\tau}\right)^2+\pi(70\tau+1)^2}\right)=$$
$$-1.70658\times 10^7\sqrt{\pi(70\tau+1)^2+118635\tau}\tau^{3/2}+$$
$$4200\sqrt{35}\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}\tau^{3/2}+$$
$$8.60302\times 10^6\sqrt{\pi(140\tau+1)^2+118635\tau}\tau^{3/2}-2.89498\times 10^7\tau^{3/2}-$$
$$1.21486\times 10^7\sqrt{\pi(70\tau+1)^2+118635\tau}\tau^{5/2}+8.8828\times 10^6\sqrt{\pi(140\tau+1)^2+118635\tau}\tau^{5/2}-$$
$$2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+2.24868\times 10^7\tau^2+94840.5\sqrt{\pi(70\tau+1)^2+118635\tau}\tau+$$
$$47420.2\sqrt{\pi(140\tau+1)^2+118635\tau}\tau+481860\tau+710.354\sqrt{\tau}+$$
$$820.213\sqrt{\tau}\sqrt{\pi(70\tau+1)^2+118635\tau}+677.432\sqrt{\pi(70\tau+1)^2+118635\tau}-$$
$$1011.27\sqrt{\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}-$$
$$20\sqrt{35}\sqrt{\tau}\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}+$$
$$200\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}+$$
$$677.432\sqrt{\pi(140\tau+1)^2+118635\tau}+2294.57=$$
$$-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+$$
$$\left(-1.70658\times 10^7\tau^{3/2}-1.21486\times 10^7\tau^{5/2}+94840.5\tau+820.213\sqrt{\tau}+677.432\right)$$
$$\sqrt{\pi(70\tau+1)^2+118635\tau}+$$
$$\left(8.60302\times 10^6\tau^{3/2}+8.8828\times 10^6\tau^{5/2}+47420.2\tau-1011.27\sqrt{\tau}+677.432\right)$$
$$\sqrt{\pi(140\tau+1)^2+118635\tau}+$$
$$\left(4200\sqrt{35}\tau^{3/2}-20\sqrt{35}\sqrt{\tau}+200\right)\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}+$$

$$
\begin{aligned}
&2.24868\times 10^7\tau^2+481860.\tau+710.354\sqrt{\tau}+2294.57\ \leqslant\\
&-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+\\
&\left(-1.70658\times 10^7\tau^{3/2}-1.21486\times 10^7\tau^{5/2}+820.213\sqrt{1.25}+1.25\cdot 94840.5+677.432\right)\\
&\sqrt{\pi(70\tau+1)^2+118635\tau}+\\
&\left(8.60302\times 10^6\tau^{3/2}+8.8828\times 10^6\tau^{5/2}-1011.27\sqrt{0.8}+1.25\cdot 47420.2+677.432\right)\\
&\sqrt{\pi(140\tau+1)^2+118635\tau}+\\
&\left(4200\sqrt{35}\tau^{3/2}-20\sqrt{35}\sqrt{\tau}+200\right)\\
&\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}+\\
&2.24868\times 10^7\tau^2+710.354\sqrt{1.25}+1.25\cdot 481860+2294.57\ =\\
&-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+\\
&\left(-1.70658\times 10^7\tau^{3/2}-1.21486\times 10^7\tau^{5/2}+120145.\right)\sqrt{\pi(70\tau+1)^2+118635\tau}+\\
&\left(8.60302\times 10^6\tau^{3/2}+8.8828\times 10^6\tau^{5/2}+59048.2\right)\sqrt{\pi(140\tau+1)^2+118635\tau}+\\
&\left(4200\sqrt{35}\tau^{3/2}-20\sqrt{35}\sqrt{\tau}+200\right)\sqrt{\pi(70\tau+1)^2+118635\tau}\sqrt{\pi(140\tau+1)^2+118635\tau}+\\
&2.24868\times 10^7\tau^2+605413\ =\\
&-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+\\
&\left(8.60302\times 10^6\tau^{3/2}+8.8828\times 10^6\tau^{5/2}+59048.2\right)\sqrt{19600\pi(\tau+1.94093)(\tau+0.0000262866)}+\\
&\left(-1.70658\times 10^7\tau^{3/2}-1.21486\times 10^7\tau^{5/2}+120145.\right)\sqrt{4900\pi(\tau+7.73521)(\tau+0.0000263835)}+\\
&\left(4200\sqrt{35}\tau^{3/2}-20\sqrt{35}\sqrt{\tau}+200\right)\\
&\sqrt{19600\pi(\tau+1.94093)(\tau+0.0000262866)}\sqrt{4900\pi(\tau+7.73521)(\tau+0.0000263835)}+\\
&2.24868\times 10^7\tau^2+605413\ \leqslant\\
&-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+\\
&\left(8.60302\times 10^6\tau^{3/2}+8.8828\times 10^6\tau^{5/2}+59048.2\right)\sqrt{19600\pi(\tau+1.94093)\tau}+\\
&\left(-1.70658\times 10^7\tau^{3/2}-1.21486\times 10^7\tau^{5/2}+120145.\right)\sqrt{4900\pi 1.00003(\tau+7.73521)\tau}+\\
&\left(4200\sqrt{35}\tau^{3/2}-20\sqrt{35}\sqrt{\tau}+200\right)\sqrt{19600\pi 1.00003(\tau+1.94093)\tau}\\
&\sqrt{4900\pi 1.00003(\tau+7.73521)\tau}+\\
&2.24868\times 10^7\tau^2+605413\ =\\
&-2.89498\times 10^7\tau^{3/2}-2.43651\times 10^7\tau^{5/2}-1.46191\times 10^9\tau^{7/2}+\\
&\left(-3.64296\times 10^6\tau^{3/2}+7.65021\times 10^8\tau^{5/2}+6.15772\times 10^6\tau\right)\\
&\sqrt{\tau+1.94093}\sqrt{\tau+7.73521}+2.24868\times 10^7\tau^2+\\
&\left(2.20425\times 10^9\tau^3+2.13482\times 10^9\tau^2+1.46527\times 10^7\sqrt{\tau}\right)\sqrt{\tau+1.94093}+\\
&\left(-1.5073\times 10^9\tau^3-2.11738\times 10^9\tau^2+1.49066\times 10^7\sqrt{\tau}\right)\sqrt{\tau+7.73521}+605413\ \leqslant\\
&\sqrt{1.25+1.94093}\sqrt{1.25+7.73521}\left(-3.64296\times 10^6\tau^{3/2}+7.65021\times 10^8\tau^{5/2}+6.15772\times 10^6\tau\right)+\\
&\sqrt{1.25+1.94093}\left(2.20425\times 10^9\tau^3+2.13482\times 10^9\tau^2+1.46527\times 10^7\sqrt{\tau}\right)+\\
&\sqrt{0.8+7.73521}\left(-1.5073\times 10^9\tau^3-2.11738\times 10^9\tau^2+1.49066\times 10^7\sqrt{\tau}\right)-
\end{aligned}
$$

$$
\begin{aligned}
&2.89498 \times 10^7 \tau^{3/2} - 2.43651 \times 10^7 \tau^{5/2} - 1.46191 \times 10^9 \tau^{7/2} + 2.24868 \times 10^7 \tau^2 + 605413 \ = \\
&- 4.84561 \times 10^7 \tau^{3/2} + 4.07198 \times 10^9 \tau^{5/2} - 1.46191 \times 10^9 \tau^{7/2} - \\
&4.66103 \times 10^8 \tau^3 - 2.34999 \times 10^9 \tau^2 + \\
&3.29718 \times 10^7 \tau + 6.97241 \times 10^7 \sqrt{\tau} + 605413 \ \leqslant \\
&\frac{605413 \tau^{3/2}}{0.8^{3/2}} - 4.84561 \times 10^7 \tau^{3/2} + \\
&4.07198 \times 10^9 \tau^{5/2} - 1.46191 \times 10^9 \tau^{7/2} - \\
&4.66103 \times 10^8 \tau^3 - 2.34999 \times 10^9 \tau^2 + \frac{3.29718 \times 10^7 \sqrt{\tau} \tau}{\sqrt{0.8}} + \frac{6.97241 \times 10^7 \tau \sqrt{\tau}}{0.8} \ = \\
&\tau^{3/2} \Big(-4.66103 \times 10^8 \tau^{3/2} - 1.46191 \times 10^9 \tau^2 - 2.34999 \times 10^9 \sqrt{\tau} + \\
&\ 4.07198 \times 10^9 \tau + 7.64087 \times 10^7 \Big) \ \leqslant \\
&\tau^{3/2} \left(-4.66103 \times 10^8 \tau^{3/2} - 1.46191 \times 10^9 \tau^2 + \frac{7.64087 \times 10^7 \sqrt{\tau}}{\sqrt{0.8}} - \right. \\
&\left. \ 2.34999 \times 10^9 \sqrt{\tau} + 4.07198 \times 10^9 \tau \right) \ = \\
&\tau^2 \left(-1.46191 \times 10^9 \tau^{3/2} + 4.07198 \times 10^9 \sqrt{\tau} - 4.66103 \times 10^8 \tau - 2.26457 \times 10^9 \right) \ \leqslant \\
&\left(-2.26457 \times 10^9 + 4.07198 \times 10^9 \sqrt{0.8} - 4.66103 \times 10^8 0.8 - 1.46191 \times 10^9 0.8^{3/2} \right) \tau^2 \ = \\
&- 4.14199 \times 10^7 \tau^2 \ < \ 0 \, .
\end{aligned}
$$

First we expanded the term (multiplied it out). The we put the terms multiplied by the same square root into brackets. The next inequality sign stems from inserting the maximal value of 1.25 for τ for some positive terms and value of 0.8 for negative terms. These terms are then expanded at the $=$-sign. The next equality factors the terms under the squared root. We decreased the negative term by setting $\tau = \tau + 0.0000263835$ under the root. We increased positive terms by setting $\tau + 0.000026286 = 1.00003\tau$ and $\tau + 0.000026383 = 1.00003\tau$ under the root for positive terms. The positive terms are increase, since $\frac{0.8+0.000026383}{0.8} = 1.00003$, thus $\tau + 0.000026286 < \tau + 0.000026383 \leqslant 1.00003\tau$. For the next inequality we decreased negative terms by inserting $\tau = 0.8$ and increased positive terms by inserting $\tau = 1.25$. The next equality expands the terms. We use upper bound of 1.25 and lower bound of 0.8 to obtain terms with corresponding exponents of τ.

For the last $\leqslant$-sign we used the function

$$
-1.46191 \times 10^9 \tau^{3/2} + 4.07198 \times 10^9 \sqrt{\tau} - 4.66103 \times 10^8 \tau - 2.26457 \times 10^9 \tag{304}
$$

The derivative of this function is

$$
-2.19286 \times 10^9 \sqrt{\tau} + \frac{2.03599 \times 10^9}{\sqrt{\tau}} - 4.66103 \times 10^8 \tag{305}
$$

and the second order derivative is

$$
-\frac{1.01799 \times 10^9}{\tau^{3/2}} - \frac{1.09643 \times 10^9}{\sqrt{\tau}} \ < \ 0 \, . \tag{306}
$$

The derivative at 0.8 is smaller than zero:

$$
\begin{aligned}
&- 2.19286 \times 10^9 \sqrt{0.8} - 4.66103 \times 10^8 + \frac{2.03599 \times 10^9}{\sqrt{0.8}} = \\
&- 1.51154 \times 10^8 \ < \ 0 \, .
\end{aligned} \tag{307}
$$

Since the second order derivative is negative, the derivative decreases with increasing τ. Therefore the derivative is negative for all values of τ that we consider, that is, the function Eq. (304) is strictly monotonically decreasing. The maximum of the function Eq. (304) is therefore at 0.8. We inserted 0.8 to obtain the maximum.

Consequently, the derivative of

$$\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{308}$$

with respect to τ is smaller than zero for maximal $\nu = 0.7$.

Next, we consider the function for the largest $\nu = 0.16$ and the largest $y = \mu\omega = 0.01$ for determining the derivative with respect to τ.

The expression becomes

$$\tau\left(e^{\left(\frac{\frac{16\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{16\tau}{100}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{16\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{16\tau}{100}}}\right)-e^{\left(\frac{\frac{2\,16\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{16\tau}{100}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{2\,16\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{16\tau}{100}}}\right)\right). \tag{309}$$

The derivative with respect to τ is

$$\begin{aligned}&\left(\sqrt{\pi}\left(e^{\frac{(16\tau+1)^2}{3200\tau}}(128\tau(2\tau+25)-1)\operatorname{erfc}\left(\frac{16\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)-\right.\right. \\ &\left.\left.2e^{\frac{(32\tau+1)^2}{3200\tau}}(128\tau(8\tau+25)-1)\operatorname{erfc}\left(\frac{32\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)\right)+40\sqrt{2}(48\tau-1)\sqrt{\tau}\right) \\ &\left(3200\sqrt{\pi}\tau\right)^{-1}\ .\end{aligned} \tag{310}$$

We are considering only the numerator and use again the approximation of Ren and MacKenzie [30]. The error analysis on the whole numerator gives an approximation error $1.1 < E < 12$. Therefore we add 20 to the numerator when we use the approximation of Ren and MacKenzie [30]. We obtain the inequalities:

$$\begin{aligned}&\sqrt{\pi}\left(e^{\frac{(16\tau+1)^2}{3200\tau}}(128\tau(2\tau+25)-1)\operatorname{erfc}\left(\frac{16\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)-\right. \\ &\left.2e^{\frac{(32\tau+1)^2}{3200\tau}}(128\tau(8\tau+25)-1)\operatorname{erfc}\left(\frac{32\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)\right)+40\sqrt{2}(48\tau-1)\sqrt{\tau}\ \leqslant \\ &\sqrt{\pi}\left(\frac{2.911(128\tau(2\tau+25)-1)}{\frac{\sqrt{\pi}(2.911-1)(16\tau+1)}{40\sqrt{2}\sqrt{\tau}}+\sqrt{\pi\left(\frac{16\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)^2+2.911^2}}-\right. \\ &\left.\frac{2\cdot 2.911(128\tau(8\tau+25)-1)}{\frac{\sqrt{\pi}(2.911-1)(32\tau+1)}{40\sqrt{2}\sqrt{\tau}}+\sqrt{\pi\left(\frac{32\tau+1}{40\sqrt{2}\sqrt{\tau}}\right)^2+2.911^2}}\right) \\ &+40\sqrt{2}(48\tau-1)\sqrt{\tau}+20\ = \\ &\sqrt{\pi}\left(\frac{(128\tau(2\tau+25)-1)\left(40\sqrt{2}2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(16\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(16\tau+1)^2}}-\right. \\ &\left.\frac{2(128\tau(8\tau+25)-1)\left(40\sqrt{2}2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}}\right)+ \\ &40\sqrt{2}(48\tau-1)\sqrt{\tau}+20\ = \\ &\left(\left(40\sqrt{2}(48\tau-1)\sqrt{\tau}+20\right)\left(\sqrt{\pi}(2.911-1)(16\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(16\tau+1)^2}\right)\right. \\ &\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)++\end{aligned} \tag{311}$$

$$2.911 \cdot 40\sqrt{2}\sqrt{\pi}(128\tau(2\tau+25)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)-$$
$$2\sqrt{\pi}40\sqrt{2}2.911(128\tau(8\tau+25)-1)$$
$$\sqrt{\tau}\left(\sqrt{\pi}(2.911-1)(16\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(16\tau+1)^2}\right)\Bigg)$$
$$\Bigg(\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)$$
$$\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)\Bigg)^{-1}.$$

After applying the approximation of Ren and MacKenzie [30] and adding 20, we first factored out $40\sqrt{2}\sqrt{\tau}$. Then we brought all terms to the same denominator.

We now consider the numerator:

$$\left(40\sqrt{2}(48\tau-1)\sqrt{\tau}+20\right)\left(\sqrt{\pi}(2.911-1)(16\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(16\tau+1)^2}\right) \tag{312}$$
$$\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)+$$
$$2.911 \cdot 40\sqrt{2}\sqrt{\pi}(128\tau(2\tau+25)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(32\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(32\tau+1)^2}\right)-$$
$$2\sqrt{\pi}40\sqrt{2}2.911(128\tau(8\tau+25)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(16\tau+1)+\sqrt{\left(40\sqrt{2}2.911\sqrt{\tau}\right)^2+\pi(16\tau+1)^2}\right) =$$
$$-1.86491\times10^6\sqrt{\pi(16\tau+1)^2+27116.5\tau}\,\tau^{3/2}+$$
$$1920\sqrt{2}\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}\,\tau^{3/2}+$$
$$940121\sqrt{\pi(32\tau+1)^2+27116.5\tau}\,\tau^{3/2}-3.16357\times10^6\tau^{3/2}-$$
$$303446\sqrt{\pi(16\tau+1)^2+27116.5\tau}\,\tau^{5/2}+221873\sqrt{\pi(32\tau+1)^2+27116.5\tau}\,\tau^{5/2}-608588\tau^{5/2}-$$
$$8.34635\times10^6\tau^{7/2}+117482.\tau^2+2167.78\sqrt{\pi(16\tau+1)^2+27116.5\tau}\,\tau+$$
$$1083.89\sqrt{\pi(32\tau+1)^2+27116.5\tau}\,\tau+$$
$$11013.9\tau+339.614\sqrt{\tau}+392.137\sqrt{\tau}\sqrt{\pi(16\tau+1)^2+27116.5\tau}+$$
$$67.7432\sqrt{\pi(16\tau+1)^2+27116.5\tau}-483.478\sqrt{\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}-$$
$$40\sqrt{2}\sqrt{\tau}\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$$
$$20\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$$
$$67.7432\sqrt{\pi(32\tau+1)^2+27116.5\tau}+229.457 =$$
$$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$$
$$\left(-1.86491\times10^6\tau^{3/2}-303446\tau^{5/2}+2167.78\tau+392.137\sqrt{\tau}+67.7432\right)$$
$$\sqrt{\pi(16\tau+1)^2+27116.5\tau}+$$
$$\left(940121\tau^{3/2}+221873\tau^{5/2}+1083.89\tau-483.478\sqrt{\tau}+67.7432\right)$$

$\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$\left(1920\sqrt{2}\tau^{3/2}-40\sqrt{2}\sqrt{\tau}+20\right)\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$117482.\tau^2+11013.9\tau+339.614\sqrt{\tau}+229.457\ \leqslant$

$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$

$\left(-1.86491\times10^6\tau^{3/2}-303446\tau^{5/2}+392.137\sqrt{1.25}+1.252167.78+67.7432\right)$

$\sqrt{\pi(16\tau+1)^2+27116.5\tau}+$

$\left(940121\tau^{3/2}+221873\tau^{5/2}-483.478\sqrt{0.8}+1.251083.89+67.7432\right)$

$\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$\left(1920\sqrt{2}\tau^{3/2}-40\sqrt{2}\sqrt{\tau}+20\right)\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$117482.\tau^2+339.614\sqrt{1.25}+1.2511013.9+229.457\ =$

$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$

$\left(-1.86491\times10^6\tau^{3/2}-303446\tau^{5/2}+3215.89\right)\sqrt{\pi(16\tau+1)^2+27116.5\tau}+$

$\left(940121\tau^{3/2}+221873\tau^{5/2}+990.171\right)\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$\left(1920\sqrt{2}\tau^{3/2}-40\sqrt{2}\sqrt{\tau}+20\right)\sqrt{\pi(16\tau+1)^2+27116.5\tau}\sqrt{\pi(32\tau+1)^2+27116.5\tau}+$

$117482\tau^2+14376.6\ =$

$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$

$\left(940121\tau^{3/2}+221873\tau^{5/2}+990.171\right)\sqrt{1024\pi(\tau+8.49155)(\tau+0.000115004)}+$

$\left(-1.86491\times10^6\tau^{3/2}-303446\tau^{5/2}+3215.89\right)\sqrt{256\pi(\tau+33.8415)(\tau+0.000115428)}+$

$\left(1920\sqrt{2}\tau^{3/2}-40\sqrt{2}\sqrt{\tau}+20\right)\sqrt{1024\pi(\tau+8.49155)(\tau+0.000115004)}$

$\sqrt{256\pi(\tau+33.8415)(\tau+0.000115428)}+$

$117482.\tau^2+14376.6\ \leqslant$

$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$

$\left(940121\tau^{3/2}+221873\tau^{5/2}+990.171\right)\sqrt{1024\pi1.00014(\tau+8.49155)\tau}+$

$\left(1920\sqrt{2}\tau^{3/2}-40\sqrt{2}\sqrt{\tau}+20\right)\sqrt{256\pi1.00014(\tau+33.8415)\tau}\sqrt{1024\pi1.00014(\tau+8.49155)\tau}+$

$\left(-1.86491\times10^6\tau^{3/2}-303446\tau^{5/2}+3215.89\right)\sqrt{256\pi(\tau+33.8415)\tau}+$

$117482.\tau^2+14376.6\ =$

$-3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+$

$\left(-91003\tau^{3/2}+4.36814\times10^6\tau^{5/2}+32174.4\tau\right)\sqrt{\tau+8.49155}\sqrt{\tau+33.8415}+117482.\tau^2+$

$\left(1.25852\times10^7\tau^3+5.33261\times10^7\tau^2+56165.1\sqrt{\tau}\right)\sqrt{\tau+8.49155}+$

$\left(-8.60549\times10^6\tau^3-5.28876\times10^7\tau^2+91200.4\sqrt{\tau}\right)\sqrt{\tau+33.8415}+14376.6\ \leqslant$

$\sqrt{1.25+8.49155}\sqrt{1.25+33.8415}\left(-91003\tau^{3/2}+4.36814\times10^6\tau^{5/2}+32174.4\tau\right)+$

$\sqrt{1.25+8.49155}\left(1.25852\times10^7\tau^3+5.33261\times10^7\tau^2+56165.1\sqrt{\tau}\right)+$

$\sqrt{0.8+33.8415}\left(-8.60549\times10^6\tau^3-5.28876\times10^7\tau^2+91200.4\sqrt{\tau}\right)-$

$3.16357\times10^6\tau^{3/2}-608588\tau^{5/2}-8.34635\times10^6\tau^{7/2}+117482.\tau^2+14376.6\ =$

$$
\begin{aligned}
&-4.84613\times 10^6\tau^{3/2}+8.01543\times 10^7\tau^{5/2}-8.34635\times 10^6\tau^{7/2}-\\
&1.13691\times 10^7\tau^3-1.44725\times 10^8\tau^2+\\
&594875.\tau+712078.\sqrt{\tau}+14376.6\ \leqslant\\
&\frac{14376.6\tau^{3/2}}{0.8^{3/2}}-4.84613\times 10^6\tau^{3/2}+\\
&8.01543\times 10^7\tau^{5/2}-8.34635\times 10^6\tau^{7/2}-\\
&1.13691\times 10^7\tau^3-1.44725\times 10^8\tau^2+\frac{594875.\sqrt{\tau}\tau}{\sqrt{0.8}}+\frac{712078.\tau\sqrt{\tau}}{0.8}\ =\\
&-3.1311\cdot 10^6\tau^{3/2}-1.44725\cdot 10^8\tau^2+8.01543\cdot 10^7\tau^{5/2}-1.13691\cdot 10^7\tau^3-\\
&8.34635\cdot 10^6\tau^{7/2}\ \leqslant\\
&-3.1311\times 10^6\tau^{3/2}+\frac{8.01543\times 10^7\sqrt{1.25}\tau^{5/2}}{\sqrt{\tau}}-\\
&8.34635\times 10^6\tau^{7/2}-1.13691\times 10^7\tau^3-1.44725\times 10^8\tau^2\ =\\
&-3.1311\times 10^6\tau^{3/2}-8.34635\times 10^6\tau^{7/2}-1.13691\times 10^7\tau^3-5.51094\times 10^7\tau^2 2\ <\ 0\,.
\end{aligned}
$$

First we expanded the term (multiplied it out). The we put the terms multiplied by the same square root into brackets. The next inequality sign stems from inserting the maximal value of 1.25 for τ for some positive terms and value of 0.8 for negative terms. These terms are then expanded at the $=$-sign. The next equality factors the terms under the squared root. We decreased the negative term by setting $\tau=\tau+0.00011542$ under the root. We increased positive terms by setting $\tau+0.00011542=1.00014\tau$ and $\tau+0.000115004=1.00014\tau$ under the root for positive terms. The positive terms are increase, since $\frac{0.8+0.00011542}{0.8}<1.000142$, thus $\tau+0.000115004<\tau+0.00011542\leqslant 1.00014\tau$. For the next inequality we decreased negative terms by inserting $\tau=0.8$ and increased positive terms by inserting $\tau=1.25$. The next equality expands the terms. We use upper bound of 1.25 and lower bound of 0.8 to obtain terms with corresponding exponents of τ.

Consequently, the derivative of

$$
\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{313}
$$

with respect to τ is smaller than zero for maximal $\nu=0.16$.

Next, we consider the function for the largest $\nu=0.24$ and the largest $y=\mu\omega=0.01$ for determining the derivative with respect to τ. However we assume $0.9\leqslant\tau$, in order to restrict the domain of τ.

The expression becomes

$$
\tau\left(e^{\left(\frac{\frac{24\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{24\tau}{100}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{24\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{24\tau}{100}}}\right)-e^{\left(\frac{\frac{2\,24\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{24\tau}{100}}}\right)^2}\operatorname{erfc}\left(\frac{\frac{2\,24\tau}{100}+\frac{1}{100}}{\sqrt{2}\sqrt{\frac{24\tau}{100}}}\right)\right)\,. \tag{314}
$$

The derivative with respect to τ is

$$
\begin{aligned}
&\left(\sqrt{\pi}\left(e^{\frac{(24\tau+1)^2}{4800\tau}}(192\tau(3\tau+25)-1)\operatorname{erfc}\left(\frac{24\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)-\right.\right. && (315)\\
&\left.\left.2e^{\frac{(48\tau+1)^2}{4800\tau}}(192\tau(12\tau+25)-1)\operatorname{erfc}\left(\frac{48\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)\right)+40\sqrt{3}(72\tau-1)\sqrt{\tau}\right)\\
&\left(4800\sqrt{\pi}\tau\right)^{-1}\,.
\end{aligned}
$$

We are considering only the numerator and use again the approximation of Ren and MacKenzie [30]. The error analysis on the whole numerator gives an approximation error $14<E<32$. Therefore we add 32 to the numerator when we use the approximation of Ren and MacKenzie [30]. We obtain the inequalities:

$$
\sqrt{\pi}\left(e^{\frac{(24\tau+1)^2}{4800\tau}}(192\tau(3\tau+25)-1)\operatorname{erfc}\left(\frac{24\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)\ -\right. \tag{316}
$$

$$
\begin{aligned}
&2e^{\frac{(48\tau+1)^2}{4800\tau}}(192\tau(12\tau+25)-1)\operatorname{erfc}\left(\frac{48\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)\Bigg)+40\sqrt{3}(72\tau-1)\sqrt{\tau} \leqslant \\
&\sqrt{\pi}\left(\frac{2.911(192\tau(3\tau+25)-1)}{\frac{\sqrt{\pi}(2.911-1)(24\tau+1)}{40\sqrt{3}\sqrt{\tau}}+\sqrt{\pi\left(\frac{24\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)^2+2.911^2}}-\right. \\
&\left.\frac{2\cdot 2.911(192\tau(12\tau+25)-1)}{\frac{\sqrt{\pi}(2.911-1)(48\tau+1)}{40\sqrt{3}\sqrt{\tau}}+\sqrt{\pi\left(\frac{48\tau+1}{40\sqrt{3}\sqrt{\tau}}\right)^2+2.911^2}}\right)+ \\
&40\sqrt{3}(72\tau-1)\sqrt{\tau}+32 \;= \\
&\sqrt{\pi}\left(\frac{(192\tau(3\tau+25)-1)\left(40\sqrt{3}2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}}-\right. \\
&\left.\frac{2(192\tau(12\tau+25)-1)\left(40\sqrt{3}2.911\sqrt{\tau}\right)}{\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}}\right)+ \\
&40\sqrt{3}(72\tau-1)\sqrt{\tau}+32 \;= \\
&\Bigg(\left(40\sqrt{3}(72\tau-1)\sqrt{\tau}+32\right)\left(\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}\right) \\
&\left(\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}\right)+ \\
&2.911\cdot 40\sqrt{3}\sqrt{\pi}(192\tau(3\tau+25)-1)\sqrt{\tau} \\
&\left(\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}\right)- \\
&2\sqrt{\pi}40\sqrt{3}2.911(192\tau(12\tau+25)-1) \\
&\sqrt{\tau}\left(\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}\right)\Bigg) \\
&\Bigg(\left(\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}\right) \\
&\left(\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}\right)\Bigg)^{-1}.
\end{aligned}
$$

After applying the approximation of Ren and MacKenzie [30] and adding 200, we first factored out $40\sqrt{3}\sqrt{\tau}$. Then we brought all terms to the same denominator.

We now consider the numerator:

$$
\begin{aligned}
&\left(40\sqrt{3}(72\tau-1)\sqrt{\tau}+32\right)\left(\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}\right) \qquad (317)\\
&\left(\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}\right)\;+ \\
&2.911\cdot 40\sqrt{3}\sqrt{\pi}(192\tau(3\tau+25)-1)\sqrt{\tau}
\end{aligned}
$$

$$\left(\sqrt{\pi}(2.911-1)(48\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(48\tau+1)^2}\right)-$$
$$2\sqrt{\pi}40\sqrt{3}2.911(192\tau(12\tau+25)-1)\sqrt{\tau}$$
$$\left(\sqrt{\pi}(2.911-1)(24\tau+1)+\sqrt{\left(40\sqrt{3}2.911\sqrt{\tau}\right)^2+\pi(24\tau+1)^2}\right)\ =$$
$$-3.42607\times10^6\sqrt{\pi(24\tau+1)^2+40674.8\tau}\tau^{3/2}+$$
$$2880\sqrt{3}\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}\tau^{3/2}+$$
$$1.72711\times10^6\sqrt{\pi(48\tau+1)^2+40674.8\tau}\tau^{3/2}-5.81185\times10^6\tau^{3/2}-$$
$$836198\sqrt{\pi(24\tau+1)^2+40674.8\tau}\tau^{5/2}+611410\sqrt{\pi(48\tau+1)^2+40674.8\tau}\tau^{5/2}-$$
$$1.67707\times10^6\tau^{5/2}-$$
$$3.44998\times10^7\tau^{7/2}+422935.\tau^2+5202.68\sqrt{\pi(24\tau+1)^2+40674.8\tau}\tau+$$
$$2601.34\sqrt{\pi(48\tau+1)^2+40674.8\tau}\tau+$$
$$26433.4\tau+415.94\sqrt{\tau}+480.268\sqrt{\tau}\sqrt{\pi(24\tau+1)^2+40674.8\tau}+$$
$$108.389\sqrt{\pi(24\tau+1)^2+40674.8\tau}-592.138\sqrt{\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}-$$
$$40\sqrt{3}\sqrt{\tau}\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$32\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$108.389\sqrt{\pi(48\tau+1)^2+40674.8\tau}+367.131\ =$$
$$-5.81185\times10^6\tau^{3/2}-1.67707\times10^6\tau^{5/2}-3.44998\times10^7\tau^{7/2}+$$
$$\left(-3.42607\times10^6\tau^{3/2}-836198\tau^{5/2}+5202.68\tau+480.268\sqrt{\tau}+108.389\right)$$
$$\sqrt{\pi(24\tau+1)^2+40674.8\tau}+$$
$$\left(1.72711\times10^6\tau^{3/2}+611410\tau^{5/2}+2601.34\tau-592.138\sqrt{\tau}+108.389\right)$$
$$\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$\left(2880\sqrt{3}\tau^{3/2}-40\sqrt{3}\sqrt{\tau}+32\right)\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$422935.\tau^2+26433.4\tau+415.94\sqrt{\tau}+367.131\ \leqslant$$
$$-5.81185\times10^6\tau^{3/2}-1.67707\times10^6\tau^{5/2}-3.44998\times10^7\tau^{7/2}+$$
$$\left(-3.42607\times10^6\tau^{3/2}-836198\tau^{5/2}+480.268\sqrt{1.25}+1.255202.68+108.389\right)$$
$$\sqrt{\pi(24\tau+1)^2+40674.8\tau}+$$
$$\left(1.72711\times10^6\tau^{3/2}+611410\tau^{5/2}-592.138\sqrt{0.9}+1.252601.34+108.389\right)$$
$$\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$\left(2880\sqrt{3}\tau^{3/2}-40\sqrt{3}\sqrt{\tau}+32\right)\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$422935\tau^2+415.94\sqrt{1.25}+1.2526433.4+367.131\ =$$
$$-5.81185\times10^6\tau^{3/2}-1.67707\times10^6\tau^{5/2}-3.44998\times10^7\tau^{7/2}+$$
$$\left(-3.42607\times10^6\tau^{3/2}-836198\tau^{5/2}+7148.69\right)\sqrt{\pi(24\tau+1)^2+40674.8\tau}+$$
$$\left(1.72711\times10^6\tau^{3/2}+611410\tau^{5/2}+2798.31\right)\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$\left(2880\sqrt{3}\tau^{3/2}-40\sqrt{3}\sqrt{\tau}+32\right)\sqrt{\pi(24\tau+1)^2+40674.8\tau}\sqrt{\pi(48\tau+1)^2+40674.8\tau}+$$
$$422935\tau^2+33874\ =$$

$$
\begin{aligned}
&-5.81185\times 10^{6}\tau^{3/2}-1.67707\times 10^{6}\tau^{5/2}-3.44998\times 10^{7}\tau^{7/2}+\\
&\left(1.72711\times 10^{6}\tau^{3/2}+611410\tau^{5/2}+2798.31\right)\sqrt{2304\pi(\tau+5.66103)(\tau+0.0000766694)}+\\
&\left(-3.42607\times 10^{6}\tau^{3/2}-836198\tau^{5/2}+7148.69\right)\sqrt{576\pi(\tau+22.561)(\tau+0.0000769518)}+\\
&\left(2880\sqrt{3}\tau^{3/2}-40\sqrt{3}\sqrt{\tau}+32\right)\sqrt{2304\pi(\tau+5.66103)(\tau+0.0000766694)}\\
&\sqrt{576\pi(\tau+22.561)(\tau+0.0000769518)}+\\
&422935\tau^{2}+33874\ \leqslant\\
&-5.8118510^{6}\tau^{3/2}-1.67707\times 10^{6}\tau^{5/2}-3.44998\times 10^{7}\tau^{7/2}+\\
&\left(1.72711\times 10^{6}\tau^{3/2}+611410\tau^{5/2}+2798.31\right)\sqrt{2304\pi 1.0001(\tau+5.66103)\tau}+\\
&\left(2880\sqrt{3}\tau^{3/2}-40\sqrt{3}\sqrt{\tau}+32\right)\sqrt{2304\pi 1.0001(\tau+5.66103)\tau}\sqrt{576\pi 1.0001(\tau+22.561)\tau}+\\
&\left(-3.42607\times 10^{6}\tau^{3/2}-836198\tau^{5/2}+7148.69\right)\\
&\sqrt{576\pi(\tau+22.561)\tau}+\\
&422935\tau^{2}+33874.\ =\\
&-5.8118510^{6}\tau^{3/2}-1.67707\times 10^{6}\tau^{5/2}-3.44998\times 10^{7}\tau^{7/2}+\\
&\left(-250764.\tau^{3/2}+1.8055\times 10^{7}\tau^{5/2}+115823.\tau\right)\\
&\sqrt{\tau+5.66103}\sqrt{\tau+22.561}+422935.\tau^{2}+\\
&\left(5.20199\times 10^{7}\tau^{3}+1.46946\times 10^{8}\tau^{2}+238086.\sqrt{\tau}\right)\sqrt{\tau+5.66103}+\\
&\left(-3.55709\times 10^{7}\tau^{3}-1.45741\times 10^{8}\tau^{2}+304097.\sqrt{\tau}\right)\sqrt{\tau+22.561}+33874.\ \leqslant\\
&\sqrt{1.25+5.66103}\sqrt{1.25+22.561}\left(-250764.\tau^{3/2}+1.8055\times 10^{7}\tau^{5/2}+115823.\tau\right)+\\
&\sqrt{1.25+5.66103}\left(5.20199\times 10^{7}\tau^{3}+1.46946\times 10^{8}\tau^{2}+238086.\sqrt{\tau}\right)+\\
&\sqrt{0.9+22.561}\left(-3.55709\times 10^{7}\tau^{3}-1.45741\times 10^{8}\tau^{2}+304097.\sqrt{\tau}\right)-\\
&5.8118510^{6}\tau^{3/2}-1.67707\times 10^{6}\tau^{5/2}-3.44998\times 10^{7}\tau^{7/2}+422935.\tau^{2}+33874.\ \leqslant\\
&\frac{33874.\tau^{3/2}}{0.9^{3/2}}-9.02866\times 10^{6}\tau^{3/2}+2.29933\times 10^{8}\tau^{5/2}-3.44998\times 10^{7}\tau^{7/2}-\\
&3.5539\times 10^{7}\tau^{3}-3.19193\times 10^{8}\tau^{2}+\frac{1.48578\times 10^{6}\sqrt{\tau}\tau}{\sqrt{0.9}}+\frac{2.09884\times 10^{6}\tau\sqrt{\tau}}{0.9}\ =\\
&-5.09079\times 10^{6}\tau^{3/2}+2.29933\times 10^{8}\tau^{5/2}-\\
&3.44998\times 10^{7}\tau^{7/2}-3.5539\times 10^{7}\tau^{3}-3.19193\times 10^{8}\tau^{2}\ \leqslant\\
&-5.09079\times 10^{6}\tau^{3/2}+\frac{2.29933\times 10^{8}\sqrt{1.25}\tau^{5/2}}{\sqrt{\tau}}-3.44998\times 10^{7}\tau^{7/2}-\\
&3.5539\times 10^{7}\tau^{3}-3.19193\times 10^{8}\tau^{2}\ =\\
&-5.09079\times 10^{6}\tau^{3/2}-3.44998\times 10^{7}\tau^{7/2}-3.5539\times 10^{7}\tau^{3}-6.21197\times 10^{7}\tau^{2}\ <\ 0\,.
\end{aligned}
$$

First we expanded the term (multiplied it out). The we put the terms multiplied by the same square root into brackets. The next inequality sign stems from inserting the maximal value of 1.25 for τ for some positive terms and value of 0.9 for negative terms. These terms are then expanded at the =-sign. The next equality factors the terms under the squared root. We decreased the negative term by setting $\tau = \tau + 0.0000769518$ under the root. We increased positive terms by setting $\tau + 0.0000769518 = 1.0000962\tau$ and $\tau + 0.0000766694 = 1.0000962\tau$ under the root for positive terms. The positive terms are increase, since $\frac{0.8+0.0000769518}{0.8} < 1.0000962$, thus $\tau + 0.0000766694 < \tau + 0.0000769518 \leqslant 1.0000962\tau$. For the next inequality we decreased negative terms by inserting $\tau = 0.9$ and increased positive terms by inserting $\tau = 1.25$. The next equality expands the terms. We use upper bound of 1.25 and lower bound of 0.9 to obtain terms with corresponding exponents of τ.

Consequently, the derivative of

$$\tau\left(e^{\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)-2e^{\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)^2}\operatorname{erfc}\left(\frac{\mu\omega+2\nu\tau}{\sqrt{2}\sqrt{\nu\tau}}\right)\right) \tag{318}$$

with respect to τ is smaller than zero for maximal $\nu = 0.24$ and the domain $0.9 \leqslant \tau \leqslant 1.25$. □

Lemma 47. *In the domain* $-0.01 \leqslant y \leqslant 0.01$ *and* $0.64 \leqslant x \leqslant 1.875$*, the function* $f(x,y) = e^{\frac{1}{2}(2y+x)}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2x}}\right)$ *has a global maximum at* $y = 0.64$ *and* $x = -0.01$ *and a global minimum at* $y = 1.875$ *and* $x = 0.01$.

Proof. $f(x,y) = e^{\frac{1}{2}(2y+x)}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2x}}\right)$ is strictly monotonically decreasing in x, since its derivative with respect to x is negative:

$$\begin{aligned}
&\frac{e^{-\frac{y^2}{2x}}\left(\sqrt{\pi}x^{3/2}e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)+\sqrt{2}(y-x)\right)}{2\sqrt{\pi}x^{3/2}} < 0\\
&\iff \sqrt{\pi}x^{3/2}e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)+\sqrt{2}(y-x) < 0\\
&\sqrt{\pi}x^{3/2}e^{\frac{(x+y)^2}{2x}}\operatorname{erfc}\left(\frac{x+y}{\sqrt{2}\sqrt{x}}\right)+\sqrt{2}(y-x) \leqslant\\
&\frac{2x^{3/2}}{\frac{x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\frac{(x+y)^2}{2x}+\frac{4}{\pi}}}+y\sqrt{2}-x\sqrt{2} \leqslant\\
&\frac{2\cdot 0.64^{3/2}}{\frac{0.01+0.64}{\sqrt{2}\sqrt{0.64}}+\sqrt{\frac{(0.01+0.64)^2}{2\cdot 0.64}+\frac{4}{\pi}}}+0.01\sqrt{2}-0.64\sqrt{2} = -0.334658 < 0.
\end{aligned} \tag{319}$$

The two last inqualities come from applying Abramowitz bounds 22 and from the fact that the expression $\frac{2x^{3/2}}{\frac{x+y}{\sqrt{2}\sqrt{x}}+\sqrt{\frac{(x+y)^2}{2x}+\frac{4}{\pi}}}+y\sqrt{2}-x\sqrt{2}$ does not change monotonicity in the domain and hence the maximum must be found at the border. For $x = 0.64$ that maximizes the function $f(x,y)$ is monotonically in y, because its derivative w.r.t. y at $x = 0.64$ is

$$\begin{aligned}
&e^{y}\left(1.37713\operatorname{erfc}(0.883883y+0.565685)-1.37349e^{-0.78125(y+0.64)^2}\right) < 0\\
&\iff \left(1.37713\operatorname{erfc}(0.883883y+0.565685)-1.37349e^{-0.78125(y+0.64)^2}\right) < 0\\
&\left(1.37713\operatorname{erfc}(0.883883y+0.565685)-1.37349e^{-0.78125(y+0.64)^2}\right) \leqslant\\
&\left(1.37713\operatorname{erfc}(0.883883\cdot -0.01+0.565685)-1.37349e^{-0.78125(0.01+0.64)^2}\right) =\\
&0.5935272325870631-0.987354705867739 < 0.
\end{aligned} \tag{320}$$

Therefore, the values $y = 0.64$ and $x = -0.01$ give a global maximum of the function $f(x,y)$ in the domain $-0.01 \leqslant y \leqslant 0.01$ and $0.64 \leqslant x \leqslant 1.875$ and the values $y = 1.875$ and $x = 0.01$ give the global minimum. □

A4 Additional information on experiments

In this section, we report the hyperparameters that were considered for each method and data set and give details on the processing of the data sets.

A4.1 121 UCI Machine Learning Repository data sets: Hyperparameters

For the UCI data sets, the best hyperparameter setting was determined by a grid-search over all hyperparameter combinations using 15% of the training data as validation set. The early stopping parameter was determined on the smoothed learning curves of 100 epochs of the validation set. Smoothing was done using moving averages of 10 consecutive values. We tested "rectangular" and "conic" layers – rectangular layers have constant number of hidden units in each layer, conic layers start with the given number of hidden units in the first layer and then decrease the number of hidden units to the size of the output layer according to the geometric progession. If multiple hyperparameters provided identical performance on the validation set, we preferred settings with a higher number of layers, lower learning rates and higher dropout rates. All methods had the chance to adjust their hyperparameters to the data set at hand.

Table A4: Hyperparameters considered for self-normalizing networks in the UCI data sets.

Hyperparameter	Considered values
Number of hidden units	{1024, 512, 256}
Number of hidden layers	{2, 3, 4, 8, 16, 32}
Learning rate	{0.01, 0.1, 1}
Dropout rate	{0.05, 0}
Layer form	{rectangular, conic}

Table A5: Hyperparameters considered for ReLU networks with MS initialization in the UCI data sets.

Hyperparameter	Considered values
Number of hidden units	{1024, 512, 256}
Number of hidden layers	{2,3,4,8,16,32}
Learning rate	{0.01, 0.1, 1}
Dropout rate	{0.5, 0}
Layer form	{rectangular, conic}

Table A6: Hyperparameters considered for batch normalized networks in the UCI data sets.

Hyperparameter	Considered values
Number of hidden units	{1024, 512, 256}
Number of hidden layers	{2, 3, 4, 8, 16, 32}
Learning rate	{0.01, 0.1, 1}
Normalization	{Batchnorm}
Layer form	{rectangular, conic}

Table A7: Hyperparameters considered for weight normalized networks in the UCI data sets.

Hyperparameter	Considered values
Number of hidden units	{1024, 512, 256}
Number of hidden layers	{2, 3, 4, 8, 16, 32}
Learning rate	{0.01, 0.1, 1}
Normalization	{Weightnorm}
Layer form	{rectangular, conic}

Table A8: Hyperparameters considered for layer normalized networks in the UCI data sets.

Hyperparameter	Considered values
Number of hidden units	{1024, 512, 256}
Number of hidden layers	{2, 3, 4, 8, 16, 32}
Learning rate	{0.01, 0.1, 1}
Normalization	{Layernorm}
Layer form	{rectangular, conic}

Table A9: Hyperparameters considered for Highway networks in the UCI data sets.

Hyperparameter	Considered values
Number of hidden layers	{2, 3, 4, 8, 16, 32}
Learning rate	{0.01, 0.1, 1}
Dropout rate	{0, 0.5}

Table A10: Hyperparameters considered for Residual networks in the UCI data sets.

Hyperparameter	Considered values
Number of blocks	{2, 3, 4, 8, 16}
Number of neurons per blocks	{1024, 512, 256}
Block form	{rectangular, diavolo}
Bottleneck	{25%, 50%}
Learning rate	{0.01, 0.1, 1}

A4.2 121 UCI Machine Learning Repository data sets: detailed results

Methods compared. We used data sets and preprocessing scripts by Fernández-Delgado et al. [10] for data preparation and defining training and test sets. With several flaws in the method comparison[37] that we avoided, the authors compared 179 machine learning methods of 17 groups in their experiments. The method groups were defined by Fernández-Delgado et al. [10] as follows: Support Vector Machines, RandomForest, Multivariate adaptive regression splines (MARS), Boosting, Rule-based, logistic and multinomial regression, Discriminant Analysis (DA), Bagging, Nearest Neighbour, DecisionTree, other Ensembles, Neural Networks, Bayesian, Other Methods, generalized linear models (GLM), Partial least squares and principal component regression (PLSR), and Stacking. However, many of methods assigned to those groups were merely different implementations of the same method. Therefore, we selected one representative of each of the 17 groups for method comparison. The representative method was chosen as the group's method with the median performance across all tasks. Finally, we included 17 other machine learning methods of Fernández-Delgado et al. [10], and 6 FNNs, BatchNorm, WeightNorm, LayerNorm, Highway, Residual and MSRAinit networks, and self-normalizing neural networks (SNNs) giving a total of 24 compared methods.

Results of FNN methods for all 121 data sets. The results of the compared FNN methods can be found in Table A11.

Small and large data sets. We assigned each of the 121 UCI data sets into the group "large datasets" or "small datasets" if the had more than 1,000 data points or less, respectively. We expected that Deep Learning methods require large data sets to competitive to other machine learning methods. This resulted in 75 small and 46 large data sets.

Results. The results of the method comparison are given in Tables A12 and A13 for small and large data sets, respectively. On small data sets, SVMs performed best followed by RandomForest and SNNs. On large data sets, SNNs are the best method followed by SVMs and Random Forest.

Table A11: Comparison of FNN methods on all 121 UCI data sets.. The table reports the accuracy of FNN methods at each individual task of the 121 UCI data sets. The first column gives the name of the data set, the second the number of training data points N, the third the number of features M and the consecutive columns the accuracy values of self-normalizing networks (SNNs), ReLU networks without normalization and with MSRA initialization (MS), Highway networks (HW), Residual Networks (ResNet), networks with batch normalization (BN), weight normalization (WN), and layer normalization (LN).

dataset	N	M	SNN	MS	HW	ResNet	BN	WN	LN
abalone	4177	9	0.6657	0.6284	0.6427	0.6466	0.6303	0.6351	0.6178
acute-inflammation	120	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9000
acute-nephritis	120	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
adult	48842	15	0.8476	0.8487	0.8453	0.8484	0.8499	0.8453	0.8517
annealing	898	32	0.7600	0.7300	0.3600	0.2600	0.1200	0.6500	0.5000
arrhythmia	452	263	0.6549	0.6372	0.6283	0.6460	0.5929	0.6018	0.5752
audiology-std	196	60	0.8000	0.6800	0.7200	0.8000	0.6400	0.7200	0.8000
balance-scale	625	5	0.9231	0.9231	0.9103	0.9167	0.9231	0.9551	0.9872
balloons	16	5	1.0000	0.5000	0.2500	1.0000	1.0000	0.0000	0.7500
bank	4521	17	0.8903	0.8876	0.8885	0.8796	0.8823	0.8850	0.8920
blood	748	5	0.7701	0.7754	0.7968	0.8021	0.7647	0.7594	0.7112
breast-cancer	286	10	0.7183	0.6901	0.7465	0.7465	0.7324	0.6197	0.6620
breast-cancer-wisc	699	10	0.9714	0.9714	0.9771	0.9714	0.9829	0.9657	0.9714
breast-cancer-wisc-diag	569	31	0.9789	0.9718	0.9789	0.9507	0.9789	0.9718	0.9648
breast-cancer-wisc-prog	198	34	0.6735	0.7347	0.8367	0.8163	0.7755	0.8367	0.7959
breast-tissue	106	10	0.7308	0.4615	0.6154	0.4231	0.4615	0.5385	0.5769
car	1728	7	0.9838	0.9861	0.9560	0.9282	0.9606	0.9769	0.9907
cardiotocography-10clases	2126	22	0.8399	0.8418	0.8456	0.8173	0.7910	0.8606	0.8362
cardiotocography-3clases	2126	22	0.9153	0.8964	0.9171	0.9021	0.9096	0.8945	0.9021
chess-krvk	28056	7	0.8805	0.8606	0.5255	0.8543	0.8781	0.7673	0.8938
chess-krvkp	3196	37	0.9912	0.9900	0.9900	0.9912	0.9862	0.9912	0.9875
congressional-voting	435	17	0.6147	0.6055	0.5872	0.5963	0.5872	0.5872	0.5780
conn-bench-sonar-mines-rocks	208	61	0.7885	0.8269	0.8462	0.8077	0.7115	0.8269	0.6731
conn-bench-vowel-deterding	990	12	0.9957	0.9935	0.9784	0.9935	0.9610	0.9524	0.9935
connect-4	67557	43	0.8807	0.8831	0.8599	0.8716	0.8729	0.8833	0.8856
contrac	1473	10	0.5190	0.5136	0.5054	0.5136	0.4538	0.4755	0.4592
credit-approval	690	16	0.8430	0.8430	0.8547	0.8430	0.8721	0.9070	0.8547
cylinder-bands	512	36	0.7266	0.7656	0.7969	0.7734	0.7500	0.7578	0.7578
dermatology	366	35	0.9231	0.9121	0.9780	0.9231	0.9341	0.9451	0.9451
echocardiogram	131	11	0.8182	0.8485	0.6061	0.8485	0.8485	0.7879	0.8182
ecoli	336	8	0.8929	0.8333	0.8690	0.8214	0.8214	0.8452	0.8571
energy-y1	768	9	0.9583	0.9583	0.8802	0.8177	0.8646	0.9010	0.9479
energy-y2	768	9	0.9063	0.8958	0.9010	0.8750	0.8750	0.8906	0.8802
fertility	100	10	0.9200	0.8800	0.8800	0.8400	0.6800	0.6800	0.8800
flags	194	29	0.4583	0.4583	0.4375	0.3750	0.4167	0.4167	0.3542
glass	214	10	0.7358	0.6038	0.6415	0.6415	0.5849	0.6792	0.6981
haberman-survival	306	4	0.7368	0.7237	0.6447	0.6842	0.7368	0.7500	0.6842
hayes-roth	160	4	0.6786	0.4643	0.7857	0.7143	0.7500	0.5714	0.8929
heart-cleveland	303	14	0.6184	0.6053	0.6316	0.5658	0.5789	0.5658	0.5789
heart-hungarian	294	13	0.7945	0.8356	0.7945	0.8082	0.8493	0.7534	0.8493
heart-switzerland	123	13	0.3548	0.3871	0.5806	0.3226	0.3871	0.2581	0.5161
heart-va	200	13	0.3600	0.2600	0.4000	0.2600	0.2800	0.2200	0.2400
hepatitis	155	20	0.7692	0.7692	0.6667	0.7692	0.8718	0.8462	0.7436
hill-valley	1212	101	0.5248	0.5116	0.5000	0.5396	0.5050	0.4934	0.5050
horse-colic	368	26	0.8088	0.8529	0.7794	0.8088	0.8529	0.7059	0.7941
ilpd-indian-liver	583	10	0.6986	0.6644	0.6781	0.6712	0.5959	0.6918	0.6986

image-segmentation	2310	19	0.9114	0.9090	0.9024	0.8919	0.8481	0.8938	0.8838
ionosphere	351	34	0.8864	0.9091	0.9432	0.9545	0.9432	0.9318	0.9432
iris	150	5	0.9730	0.9189	0.8378	0.9730	0.9189	1.0000	0.9730
led-display	1000	8	0.7640	0.7200	0.7040	0.7160	0.6280	0.6920	0.6480
lenses	24	5	0.6667	1.0000	1.0000	0.6667	0.8333	0.8333	0.6667
letter	20000	17	0.9726	0.9712	0.8984	0.9762	0.9796	0.9580	0.9742
libras	360	91	0.7889	0.8667	0.8222	0.7111	0.7444	0.8000	0.8333
low-res-spect	531	101	0.8571	0.8496	0.9023	0.8647	0.8571	0.8872	0.8947
lung-cancer	32	57	0.6250	0.3750	0.1250	0.2500	0.5000	0.5000	0.2500
lymphography	148	19	0.9189	0.7297	0.7297	0.6757	0.7568	0.7568	0.7838
magic	19020	11	0.8692	0.8629	0.8673	0.8723	0.8713	0.8690	0.8620
mammographic	961	6	0.8250	0.8083	0.7917	0.7833	0.8167	0.8292	0.8208
miniboone	130064	51	0.9307	0.9250	0.9270	0.9254	0.9262	0.9272	0.9313
molec-biol-promoter	106	58	0.8462	0.7692	0.6923	0.7692	0.7692	0.6923	0.4615
molec-biol-splice	3190	61	0.9009	0.8482	0.8833	0.8557	0.8519	0.8494	0.8607
monks-1	556	7	0.7523	0.6551	0.5833	0.7546	0.9074	0.5000	0.7014
monks-2	601	7	0.5926	0.6343	0.6389	0.6273	0.3287	0.6644	0.5162
monks-3	554	7	0.6042	0.7454	0.5880	0.5833	0.5278	0.5231	0.6991
mushroom	8124	22	1.0000	1.0000	1.0000	1.0000	0.9990	0.9995	0.9995
musk-1	476	167	0.8739	0.8655	0.8992	0.8739	0.8235	0.8992	0.8992
musk-2	6598	167	0.9891	0.9945	0.9915	0.9964	0.9982	0.9927	0.9951
nursery	12960	9	0.9978	0.9988	1.0000	0.9994	0.9994	0.9966	0.9966
oocytes_merluccius_nucleus_4d	1022	42	0.8235	0.8196	0.7176	0.8000	0.8078	0.8078	0.7686
oocytes_merluccius_states_2f	1022	26	0.9529	0.9490	0.9490	0.9373	0.9333	0.9020	0.9412
oocytes_trisopterus_nucleus_2f	912	26	0.7982	0.8728	0.8289	0.7719	0.7456	0.7939	0.8202
oocytes_trisopterus_states_5b	912	33	0.9342	0.9430	0.9342	0.8947	0.8947	0.9254	0.8991
optical	5620	63	0.9711	0.9666	0.9644	0.9627	0.9716	0.9638	0.9755
ozone	2536	73	0.9700	0.9732	0.9716	0.9669	0.9669	0.9748	0.9716
page-blocks	5473	11	0.9583	0.9708	0.9656	0.9605	0.9613	0.9730	0.9708
parkinsons	195	23	0.8980	0.9184	0.8367	0.9184	0.8571	0.8163	0.8571
pendigits	10992	17	0.9706	0.9714	0.9671	0.9708	0.9734	0.9620	0.9657
pima	768	9	0.7552	0.7656	0.7188	0.7135	0.7188	0.6979	0.6927
pittsburg-bridges-MATERIAL	106	8	0.8846	0.8462	0.9231	0.9231	0.8846	0.8077	0.9231
pittsburg-bridges-REL-L	103	8	0.6923	0.7692	0.6923	0.8462	0.7692	0.6538	0.7308
pittsburg-bridges-SPAN	92	8	0.6957	0.5217	0.5652	0.5652	0.5652	0.6522	0.6087
pittsburg-bridges-T-OR-D	102	8	0.8400	0.8800	0.8800	0.8800	0.8800	0.8800	0.8800
pittsburg-bridges-TYPE	105	8	0.6538	0.6538	0.5385	0.6538	0.1154	0.4615	0.6538
planning	182	13	0.6889	0.6667	0.6000	0.7111	0.6222	0.6444	0.6889
plant-margin	1600	65	0.8125	0.8125	0.8375	0.7975	0.7600	0.8175	0.8425
plant-shape	1600	65	0.7275	0.6350	0.6325	0.5150	0.2850	0.6575	0.6775
plant-texture	1599	65	0.8125	0.7900	0.7900	0.8000	0.8200	0.8175	0.8350
post-operative	90	9	0.7273	0.7273	0.5909	0.7273	0.5909	0.5455	0.7727
primary-tumor	330	18	0.5244	0.5000	0.4512	0.3902	0.5122	0.5000	0.4512
ringnorm	7400	21	0.9751	0.9843	0.9692	0.9811	0.9843	0.9719	0.9827
seeds	210	8	0.8846	0.8654	0.9423	0.8654	0.8654	0.8846	0.8846
semeion	1593	257	0.9196	0.9296	0.9447	0.9146	0.9372	0.9322	0.9447
soybean	683	36	0.8511	0.8723	0.8617	0.8670	0.8883	0.8537	0.8484
spambase	4601	58	0.9409	0.9461	0.9435	0.9461	0.9426	0.9504	0.9513
spect	265	23	0.6398	0.6183	0.6022	0.6667	0.6344	0.6398	0.6720
spectf	267	45	0.4973	0.6043	0.8930	0.7005	0.2299	0.4545	0.5561
statlog-australian-credit	690	15	0.5988	0.6802	0.6802	0.6395	0.6802	0.6860	0.6279
statlog-german-credit	1000	25	0.7560	0.7280	0.7760	0.7720	0.7520	0.7400	0.7400

statlog-heart	270	14	0.9254	0.8358	0.7761	0.8657	0.7910	0.8657	0.7910
statlog-image	2310	19	0.9549	0.9757	0.9584	0.9584	0.9671	0.9515	0.9757
statlog-landsat	6435	37	0.9100	0.9075	0.9110	0.9055	0.9040	0.8925	0.9040
statlog-shuttle	58000	10	0.9990	0.9983	0.9977	0.9992	0.9988	0.9988	0.9987
statlog-vehicle	846	19	0.8009	0.8294	0.7962	0.7583	0.7583	0.8009	0.7915
steel-plates	1941	28	0.7835	0.7567	0.7608	0.7629	0.7031	0.7856	0.7588
synthetic-control	600	61	0.9867	0.9800	0.9867	0.9600	0.9733	0.9867	0.9733
teaching	151	6	0.5000	0.6053	0.5263	0.5526	0.5000	0.3158	0.6316
thyroid	7200	22	0.9816	0.9770	0.9708	0.9799	0.9778	0.9807	0.9752
tic-tac-toe	958	10	0.9665	0.9833	0.9749	0.9623	0.9833	0.9707	0.9791
titanic	2201	4	0.7836	0.7909	0.7927	0.7727	0.7800	0.7818	0.7891
trains	10	30	NA	NA	NA	NA	0.5000	0.5000	1.0000
twonorm	7400	21	0.9805	0.9778	0.9708	0.9735	0.9757	0.9730	0.9724
vertebral-column-2clases	310	7	0.8312	0.8701	0.8571	0.8312	0.8312	0.6623	0.8442
vertebral-column-3clases	310	7	0.8312	0.8052	0.7922	0.7532	0.7792	0.7403	0.8312
wall-following	5456	25	0.9098	0.9076	0.9230	0.9223	0.9333	0.9274	0.9128
waveform	5000	22	0.8480	0.8312	0.8320	0.8360	0.8360	0.8376	0.8448
waveform-noise	5000	41	0.8608	0.8328	0.8696	0.8584	0.8480	0.8640	0.8504
wine	178	14	0.9773	0.9318	0.9091	0.9773	0.9773	0.9773	0.9773
wine-quality-red	1599	12	0.6300	0.6250	0.5625	0.6150	0.5450	0.5575	0.6100
wine-quality-white	4898	12	0.6373	0.6479	0.5564	0.6307	0.5335	0.5482	0.6544
yeast	1484	9	0.6307	0.6173	0.6065	0.5499	0.4906	0.5876	0.6092
zoo	101	17	0.9200	1.0000	0.8800	1.0000	0.7200	0.9600	0.9600

Table A12: UCI comparison reporting the average rank of a method on 75 classification task of the UCI machine learning repository with less than 1000 data points. For each dataset, the 24 compared methods, were ranked by their accuracy and the ranks were averaged across the tasks. The first column gives the method group, the second the method, the third the average rank , and the last the p-value of a paired Wilcoxon test whether the difference to the best performing method is significant. SNNs are ranked third having been outperformed by Random Forests and SVMs.

methodGroup	method	avg. rank	p-value
SVM	LibSVM_weka	9.3	
RandomForest	RRFglobal_caret	9.6	2.5e-01
SNN	SNN	9.6	3.8e-01
LMR	SimpleLogistic_weka	9.9	1.5e-01
NeuralNetworks	lvq_caret	10.1	1.0e-01
MARS	gcvEarth_caret	10.7	3.6e-02
MSRAinit	MSRAinit	11.0	4.0e-02
LayerNorm	LayerNorm	11.3	7.2e-02
Highway	Highway	11.5	8.9e-03
DiscriminantAnalysis	mda_R	11.8	2.6e-03
Boosting	LogitBoost_weka	11.9	2.4e-02
Bagging	ctreeBag_R	12.1	1.8e-03
ResNet	ResNet	12.3	3.5e-03
BatchNorm	BatchNorm	12.6	4.9e-04
Rule-based	JRip_caret	12.9	1.7e-04
WeightNorm	WeightNorm	13.0	8.3e-05
DecisionTree	rpart2_caret	13.6	7.0e-04
OtherEnsembles	Dagging_weka	13.9	3.0e-05
Nearest Neighbour	NNge_weka	14.0	7.7e-04
OtherMethods	pam_caret	14.2	1.5e-04
PLSR	simpls_R	14.3	4.6e-05
Bayesian	NaiveBayes_weka	14.6	1.2e-04
GLM	bayesglm_caret	15.0	1.6e-06
Stacking	Stacking_weka	20.9	2.2e-12

Table A13: UCI comparison reporting the average rank of a method on 46 classification task of the UCI machine learning repository with more than 1000 data points. For each dataset, the 24 compared methods, were ranked by their accuracy and the ranks were averaged across the tasks. The first column gives the method group, the second the method, the third the average rank , and the last the p-value of a paired Wilcoxon test whether the difference to the best performing method is significant. SNNs are ranked first having outperformed diverse machine learning methods and other FNNs.

methodGroup	method	avg. rank	p-value
SNN	SNN	5.8	
SVM	LibSVM_weka	6.1	5.8e-01
RandomForest	RRFglobal_caret	6.6	2.1e-01
MSRAinit	MSRAinit	7.1	4.5e-03
LayerNorm	LayerNorm	7.2	7.1e-02
Highway	Highway	7.9	1.7e-03
ResNet	ResNet	8.4	1.7e-04
WeightNorm	WeightNorm	8.7	5.5e-04
BatchNorm	BatchNorm	9.7	1.8e-04
MARS	gcvEarth_caret	9.9	8.2e-05
Boosting	LogitBoost_weka	12.1	2.2e-07
LMR	SimpleLogistic_weka	12.4	3.8e-09
Rule-based	JRip_caret	12.4	9.0e-08
Bagging	ctreeBag_R	13.5	1.6e-05
DiscriminantAnalysis	mda_R	13.9	1.4e-10
Nearest Neighbour	NNge_weka	14.1	1.6e-10
DecisionTree	rpart2_caret	15.5	2.3e-08
OtherEnsembles	Dagging_weka	16.1	4.4e-12
NeuralNetworks	lvq_caret	16.3	1.6e-12
Bayesian	NaiveBayes_weka	17.9	1.6e-12
OtherMethods	pam_caret	18.3	2.8e-14
GLM	bayesglm_caret	18.7	1.5e-11
PLSR	simpls_R	19.0	3.4e-11
Stacking	Stacking_weka	22.5	2.8e-14

A4.3 Tox21 challenge data set: Hyperparameters

For the Tox21 data set, the best hyperparameter setting was determined by a grid-search over all hyperparameter combinations using the validation set defined by the challenge winners [28]. The hyperparameter space was chosen to be similar to the hyperparameters that were tested by Mayr et al. [28]. The early stopping parameter was determined on the smoothed learning curves of 100 epochs of the validation set. Smoothing was done using moving averages of 10 consecutive values. We tested "rectangular" and "conic" layers – rectangular layers have constant number of hidden units in each layer, conic layers start with the given number of hidden units in the first layer and then decrease the number of hidden units to the size of the output layer according to the geometric progession. All methods had the chance to adjust their hyperparameters to the data set at hand.

Table A14: Hyperparameters considered for self-normalizing networks in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden units	{1024, 2048}
Number of hidden layers	{2,3,4,6,8,16,32}
Learning rate	{0.01, 0.05, 0.1}
Dropout rate	{0.05, 0.10}
Layer form	{rectangular, conic}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A15: Hyperparameters considered for ReLU networks with MS initialization in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden units	{1024, 2048}
Number of hidden layers	{2,3,4,6,8,16,32}
Learning rate	{0.01, 0.05, 0.1}
Dropout rate	{0.5, 0}
Layer form	{rectangular, conic}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A16: Hyperparameters considered for batch normalized networks in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden units	{1024, 2048}
Number of hidden layers	{2, 3, 4, 6, 8, 16, 32}
Learning rate	{0.01, 0.05, 0.1}
Normalization	{Batchnorm}
Layer form	{rectangular, conic}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A17: Hyperparameters considered for weight normalized networks in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden units	{1024, 2048}
Number of hidden layers	{2, 3, 4, 6, 8, 16, 32}
Learning rate	{0.01, 0.05, 0.1}
Normalization	{Weightnorm}
Dropout rate	{0, 0.5}
Layer form	{rectangular, conic}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A18: Hyperparameters considered for layer normalized networks in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden units	{1024, 2048}
Number of hidden layers	{2, 3, 4, 6, 8, 16, 32}
Learning rate	{0.01, 0.05, 0.1}
Normalization	{Layernorm}
Dropout rate	{0, 0.5}
Layer form	{rectangular, conic}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A19: Hyperparameters considered for Highway networks in the Tox21 data set.

Hyperparameter	Considered values
Number of hidden layers	{2, 3, 4, 6, 8, 16, 32}
Learning rate	{0.01, 0.05, 0.1}
Dropout rate	{0, 0.5}
L2 regularization parameter	{0.001,0.0001,0.00001}

Table A20: Hyperparameters considered for Residual networks in the Tox21 data set.

Hyperparameter	Considered values
Number of blocks	{2, 3, 4, 6, 8, 16}
Number of neurons per blocks	{1024, 2048}
Block form	{rectangular, diavolo}
Bottleneck	{25%, 50%}
Learning rate	{0.01, 0.05, 0.1}
L2 regularization parameter	{0.001,0.0001,0.00001}

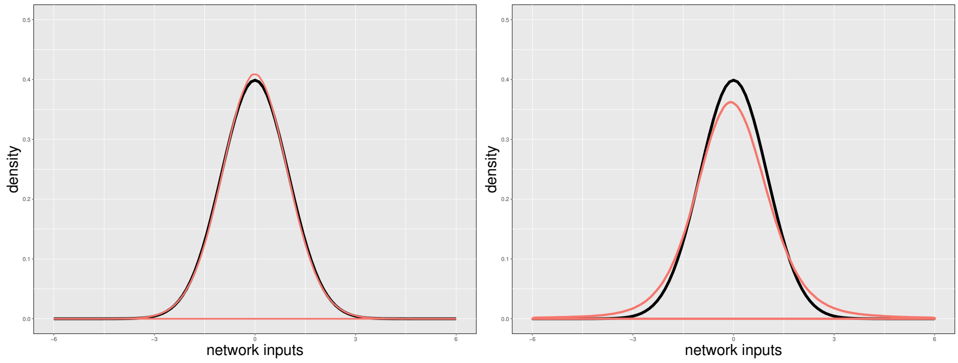

Figure A8: Distribution of network inputs of an SNN for the Tox21 data set. The plots show the distribution of network inputs z of the second layer of a typical Tox21 network. The red curves display a kernel density estimator of the network inputs and the black curve is the density of a standard normal distribution. **Left panel:** At initialization time before learning. The distribution of network inputs is close to a standard normal distribution. **Right panel:** After 40 epochs of learning. The distributions of network inputs is close to a normal distribution.

Distribution of network inputs. We empirically checked the assumption that the distribution of network inputs can well be approximated by a normal distribution. To this end, we investigated the density of the network inputs before and during learning and found that these density are close to normal distributions (see Figure A8).

A4.4 HTRU2 data set: Hyperparameters

For the HTRU2 data set, the best hyperparameter setting was determined by a grid-search over all hyperparameter combinations using one of the 9 non-testing folds as validation fold in a nested cross-validation procedure. Concretely, if M was the testing fold, we used $M - 1$ as validation fold, and for $M = 1$ we used fold 10 for validation. The early stopping parameter was determined on the smoothed learning curves of 100 epochs of the validation set. Smoothing was done using moving averages of 10 consecutive values. We tested "rectangular" and "conic" layers – rectangular layers have constant number of hidden units in each layer, conic layers start with the given number of hidden units in the first layer and then decrease the number of hidden units to the size of the output layer according to the geometric progession. All methods had the chance to adjust their hyperparameters to the data set at hand.

Table A21: Hyperparameters considered for self-normalizing networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Dropout rate	{ 0, 0.05}
Layer form	{rectangular, conic}

Table A22: Hyperparameters considered for ReLU networks with Microsoft initialization on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Dropout rate	{0, 0.5}
Layer form	{rectangular, conic}

Table A23: Hyperparameters considered for BatchNorm networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Normalization	{Batchnorm}
Layer form	{rectangular, conic}

Table A24: Hyperparameters considered for WeightNorm networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Normalization	{Weightnorm}
Layer form	{rectangular, conic}

Table A25: Hyperparameters considered for LayerNorm networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Normalization	{Layernorm}
Layer form	{rectangular, conic}

Table A26: Hyperparameters considered for Highway networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden layers	{2, 4, 8, 16, 32}
Learning rate	{0.1, 0.01, 1}
Dropout rate	{0, 0.5}

Table A27: Hyperparameters considered for Residual networks on the HTRU2 data set.

Hyperparameter	Considered values
Number of hidden units	{256, 512, 1024}
Number of residual blocks	{2, 3, 4, 8, 16}
Learning rate	{0.1, 0.01, 1}
Block form	{rectangular, diavolo}
Bottleneck	{0.25, 0.5}

A5 Other fixed points

A similar analysis with corresponding function domains can be performed for other fixed points, for example for $\mu = \tilde{\mu} = 0$ and $\nu = \tilde{\nu} = 2$, which leads to a SELU activation function with parameters $\alpha_{02} = 1.97126$ and $\lambda_{02} = 1.06071$.

A6 Bounds determined by numerical methods

In this section we report bounds on previously discussed expressions as determined by numerical methods (min and max have been computed).

$$
\begin{aligned}
0_{(\mu=0.06,\omega=0,\nu=1.35,\tau=1.12)} &< \frac{\partial \mathcal{J}_{11}}{\partial \mu} < .00182415_{(\mu=-0.1,\omega=0.1,\nu=1.47845,\tau=0.883374)} \qquad (321)\\
0.905413_{(\mu=0.1,\omega=-0.1,\nu=1.5,\tau=1.25)} &< \frac{\partial \mathcal{J}_{11}}{\partial \omega} < 1.04143_{(\mu=0.1,\omega=0.1,\nu=0.8,\tau=0.8)}\\
-0.0151177_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{11}}{\partial \nu} < 0.0151177_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)}\\
-0.015194_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{11}}{\partial \tau} < 0.015194_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)}\\
-0.0151177_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{12}}{\partial \mu} < 0.0151177_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)}\\
-0.0151177_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{12}}{\partial \omega} < 0.0151177_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)}\\
-0.00785613_{(\mu=0.1,\omega=-0.1,\nu=1.5,\tau=1.25)} &< \frac{\partial \mathcal{J}_{12}}{\partial \nu} < 0.0315805_{(\mu=0.1,\omega=0.1,\nu=0.8,\tau=0.8)}\\
0.0799824_{(\mu=0.1,\omega=-0.1,\nu=1.5,\tau=1.25)} &< \frac{\partial \mathcal{J}_{12}}{\partial \tau} < 0.110267_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=0.8)}\\
0_{(\mu=0.06,\omega=0,\nu=1.35,\tau=1.12)} &< \frac{\partial \mathcal{J}_{21}}{\partial \mu} < 0.0174802_{(\mu=0.1,\omega=0.1,\nu=0.8,\tau=0.8)}\\
0.0849308_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=0.8)} &< \frac{\partial \mathcal{J}_{21}}{\partial \omega} < 0.695766_{(\mu=0.1,\omega=0.1,\nu=1.5,\tau=1.25)}\\
-0.0600823_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{21}}{\partial \nu} < 0.0600823_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)}\\
-0.0673083_{(\mu=0.1,\omega=-0.1,\nu=1.5,\tau=0.8)} &< \frac{\partial \mathcal{J}_{21}}{\partial \tau} < 0.0673083_{(\mu=-0.1,\omega=0.1,\nu=1.5,\tau=0.8)}\\
-0.0600823_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{22}}{\partial \mu} < 0.0600823_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)}\\
-0.0600823_{(\mu=0.1,\omega=-0.1,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{22}}{\partial \omega} < 0.0600823_{(\mu=-0.1,\omega=0.1,\nu=0.8,\tau=1.25)}\\
-0.276862_{(\mu=-0.01,\omega=-0.01,\nu=0.8,\tau=1.25)} &< \frac{\partial \mathcal{J}_{22}}{\partial \nu} < -0.084813_{(\mu=-0.1,\omega=0.1,\nu=1.5,\tau=0.8)}\\
0.562302_{(\mu=0.1,\omega=-0.1,\nu=1.5,\tau=1.25)} &< \frac{\partial \mathcal{J}_{22}}{\partial \tau} < 0.664051_{(\mu=0.1,\omega=0.1,\nu=0.8,\tau=0.8)}
\end{aligned}
$$

$$
\begin{aligned}
\left|\frac{\partial \mathcal{J}_{11}}{\partial \mu}\right| &< 0.00182415(0.0031049101995398316) \qquad (322)\\
\left|\frac{\partial \mathcal{J}_{11}}{\partial \omega}\right| &< 1.04143(1.055872374194189)\\
\left|\frac{\partial \mathcal{J}_{11}}{\partial \nu}\right| &< 0.0151177(0.031242911235461816)
\end{aligned}
$$

$$\left|\frac{\partial \mathcal{J}_{11}}{\partial \tau}\right| < 0.015194(0.03749149348255419)$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \mu}\right| < 0.0151177(0.031242911235461816)$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \omega}\right| < 0.0151177(0.031242911235461816)$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \nu}\right| < 0.0315805(0.21232788238624354)$$
$$\left|\frac{\partial \mathcal{J}_{12}}{\partial \tau}\right| < 0.110267(0.2124377655377270)$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \mu}\right| < 0.0174802(0.02220441024325437)$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \omega}\right| < 0.695766(1.146955401845684)$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \nu}\right| < 0.0600823(0.14983446469110305)$$
$$\left|\frac{\partial \mathcal{J}_{21}}{\partial \tau}\right| < 0.0673083(0.17980135762932363)$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \mu}\right| < 0.0600823(0.14983446469110305)$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \omega}\right| < 0.0600823(0.14983446469110305)$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \nu}\right| < 0.562302(1.805740052651535)$$
$$\left|\frac{\partial \mathcal{J}_{22}}{\partial \tau}\right| < 0.664051(2.396685907216327)$$

A7 References

[1] Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, 10th edition.

[2] Ba, J. L., Kiros, J. R., and Hinton, G. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

[3] Bengio, Y. (2013). Deep learning of representations: Looking forward. In *Proceedings of the First International Conference on Statistical Language and Speech Processing*, pages 1–37, Berlin, Heidelberg.

[4] Blinn, J. (1996). Consider the lowly 2×2 matrix. *IEEE Computer Graphics and Applications*, pages 82–88.

[5] Bradley, R. C. (1981). Central limit theorems under weak dependence. *Journal of Multivariate Analysis*, 11(1):1–16.

[6] Cireşan, D. and Meier, U. (2015). Multi-column deep neural networks for offline handwritten chinese character classification. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.

[7] Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). *5th International Conference on Learning Representations*, arXiv:1511.07289.

[8] Dugan, P., Clark, C., LeCun, Y., and Van Parijs, S. (2016). Phase 4: Dcl system using deep learning approaches for land-based or ship-based real-time recognition and localization of marine mammals-distributed processing and big data applications. *arXiv preprint arXiv:1605.00982*.

[9] Esteva, A., Kuprel, B., Novoa, R., Ko, J., Swetter, S., Blau, H., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118.

[10] Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems. *Journal of Machine Learning Research*, 15(1):3133–3181.

[11] Goldberg, D. (1991). What every computer scientist should know about floating-point arithmetic. *ACM Comput. Surv.*, 223(1):5–48.

[12] Graves, A., Mohamed, A., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 6645–6649.

[13] Graves, A. and Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in neural information processing systems*, pages 545–552.

[14] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410.

[15] Harrison, J. (1999). A machine-checked theory of floating point arithmetic. In Bertot, Y., Dowek, G., Hirschowitz, A., Paulin, C., and Théry, L., editors, *Theorem Proving in Higher Order Logics: 12th International Conference, TPHOLs'99*, volume 1690 of *Lecture Notes in Computer Science*, pages 113–130. Springer-Verlag.

[16] He, K., Zhang, X., Ren, S., and Sun, J. (2015a). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[17] He, K., Zhang, X., Ren, S., and Sun, J. (2015b). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034.

[18] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

[19] Huval, B., Wang, T., Tandon, S., et al. (2015). An empirical evaluation of deep learning on highway driving. *arXiv preprint arXiv:1504.01716*.

[20] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 448–456.

[21] Kahan, W. (2004). A logarithm too clever by half. Technical report, University of California, Berkeley.

[22] Korolev, V. and Shevtsova, I. (2012). An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal*, 2012(2):81–105.

[23] Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

[24] LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.

[25] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[26] Loosemore, S., Stallman, R. M., McGrath, R., Oram, A., and Drepper, U. (2016). *The GNU C Library: Application Fundamentals*. GNU Press, Free Software Foundation, 51 Franklin St, Fifth Floor, Boston, MA 02110-1301, USA, 2.24 edition.

[27] Lyon, R., Stappers, B., Cooper, S., Brooke, J., and Knowles, J. (2016). Fifty years of pulsar candidate selection: From simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, 459(1):1104–1123.

[28] Mayr, A., Klambauer, G., Unterthiner, T., and Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. *Frontiers in Environmental Science*, 3:80.

[29] Muller, J.-M. (2005). On the definition of $\mathrm{ulp}(x)$. Technical Report Research report RR2005-09, Laboratoire de l'Informatique du Parallélisme.

[30] Ren, C. and MacKenzie, A. R. (2007). Closed-form approximations to the error and complementary error functions and their applications in atmospheric science. *Atmos. Sci. Let.*, pages 70–73.

[31] Sak, H., Senior, A., Rao, K., and Beaufays, F. (2015). Fast and accurate recurrent neural network acoustic models for speech recognition. *arXiv preprint arXiv:1507.06947*.

[32] Salimans, T. and Kingma, D. P. (2016). Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909.

[33] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.

[34] Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

[35] Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015). Training very deep networks. In *Advances in Neural Information Processing Systems*, pages 2377–2385.

[36] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112.

[37] Wainberg, M., Alipanahi, B., and Frey, B. J. (2016). Are random forests truly the best classifiers? *Journal of Machine Learning Research*, 17(110):1–5.

List of Figures

List of Tables

Brief index