Homework #1

John Karasev

STAT 672

January 22, 2020

1 Orthogonal projection in a hilbert space

$$\pi_V(f) \in V;$$
$$\langle f - \pi_V(f), g\rangle = 0 \ \forall g \in V;$$

1.

$$\begin{aligned}
\|f\|^2 &= \langle f, f\rangle \\
&= \langle \pi_v(f) + (f - \pi_v(f)), \pi_v(f) + (f - \pi_v(f))\rangle \\
&= \langle \pi_v(f), \pi_v(f)\rangle + 2\langle \pi_v(f), f - \pi_v(f)\rangle + \langle f - \pi_v(f), f - \pi_v(f)\rangle \\
&= \|\pi_v(f)\|^2 + \|f - \pi_v(f)\|^2 \ \textit{center term cancels since orthogonal}
\end{aligned}$$

2.

$$\begin{aligned}
\langle f, \pi_v(g)\rangle &= \langle (f - \pi_v(f)) + \pi_v(f), \pi_v(g)\rangle \\
&= \langle f - \pi_v(f), \pi_v(g)\rangle + \langle \pi_v(f), \pi_v(g)\rangle \\
&= \langle \pi_v(f), \pi_v(g)\rangle \ \textit{by orthogonality}
\end{aligned}$$

$$\begin{aligned}
\langle g, \pi_v(f)\rangle &= \langle (g - \pi_v(g)) + \pi_v(g), \pi_v(f)\rangle \\
&= \langle g - \pi_v(g), \pi_v(f)\rangle + \langle \pi_v(g), \pi_v(f)\rangle \\
&= \langle \pi_v(g), \pi_v(f)\rangle \ \textit{by orthogonality}
\end{aligned}$$

$$\langle g, \pi_v(f)\rangle = \langle f, \pi_v(g)\rangle$$

3.

2 A reproducing kernel is positive definite

1.

$$\begin{aligned}
K(x,y) &= \langle K_x(.), K_y(.)\rangle \ \textit{by reproducing property} \\
K(y,x) &= \langle K_y(.), K_x(.)\rangle \ \textit{by reproducing property} \\
K(x,y) &= \langle K_x(.), K_y(.)\rangle = \langle K_y(.), K_x(.)\rangle \ = K(y,x)
\end{aligned}$$

2.

$$\textit{for any } n \in \mathbb{N}, \ (a_1, ..., a_n) \in \mathbb{R}^n \ ,$$
$$(x_1, ..., x_n) \in \mathbb{X}^n$$

$$\begin{aligned}
\sum_{i,j=1}^{n} a_i a_j K(x_i, x_j) &= \sum_{i,j=1}^{n} a_i a_j \langle Kx_i, Kx_j\rangle_{\mathcal{H}} \qquad (1) \\
&= \left\| \sum_{i=1}^{n} a_i K x_i \right\|_{\mathcal{H}} \geq 0
\end{aligned}$$

3.

$$|\langle f, K_x\rangle| \leq \|f\|\|K_x\| \; true\ by\ Cauchy - Schwartz\ inequality$$
$$|\langle f, K_x\rangle| \leq \|f\|\sqrt{\langle K_x, K_x\rangle} \; by\ norm\ definition$$
$$|f(x)| \leq \|f\|\sqrt{K(x,x)} \; reproducing\ property$$

3 RKHS over a finite set

1.

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \boldsymbol{f}^T\boldsymbol{K}^{-1}\boldsymbol{g} \quad \boldsymbol{f}, \boldsymbol{g} \in \mathbb{R}^{2m+1}$$

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\left\{\frac{-(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right\} \quad \boldsymbol{x} \in \mathbb{R}^n$$

$$\begin{aligned} p(\boldsymbol{f}) &= Ce^{-\frac{\|\boldsymbol{f}\|_H^2}{2}} \\ &= Cexp\left\{-\frac{\boldsymbol{f}^T\boldsymbol{K}^{-1}\boldsymbol{f}}{2}\right\} \\ &= \frac{1}{(2\pi)^{\frac{2m+1}{2}}|\boldsymbol{K}|^{\frac{1}{2}}} exp\left\{-\frac{\boldsymbol{f}^T\boldsymbol{K}^{-1}\boldsymbol{f}}{2}\right\} \\ p(\boldsymbol{f}) &= \mathcal{N}(\boldsymbol{f}; \boldsymbol{0}, \boldsymbol{K}) \end{aligned}$$

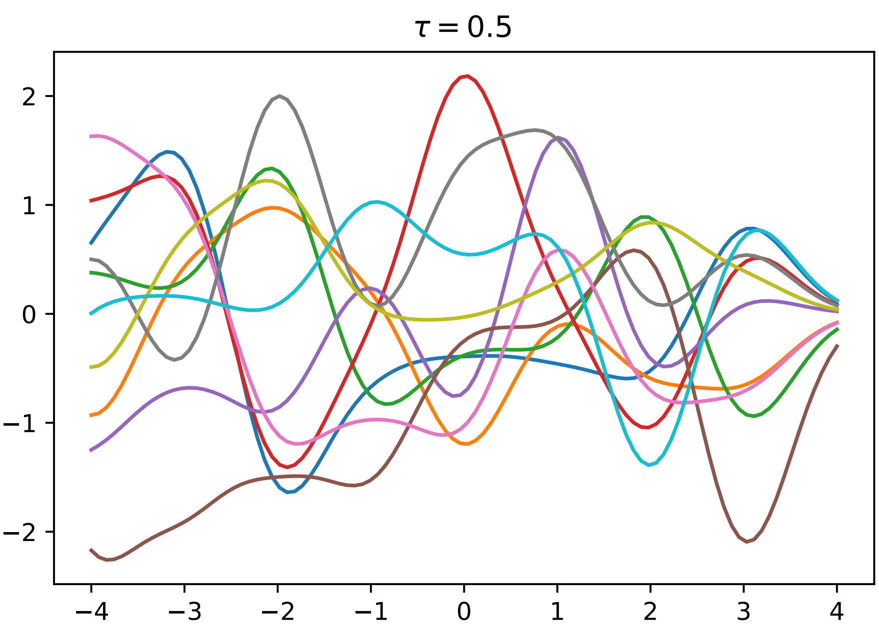
$$\therefore \; \boldsymbol{\mu} = \boldsymbol{0}$$
$$\boldsymbol{\Sigma} = \boldsymbol{K}$$

2. C is a normalizing constant so the probability distribution sums to 1

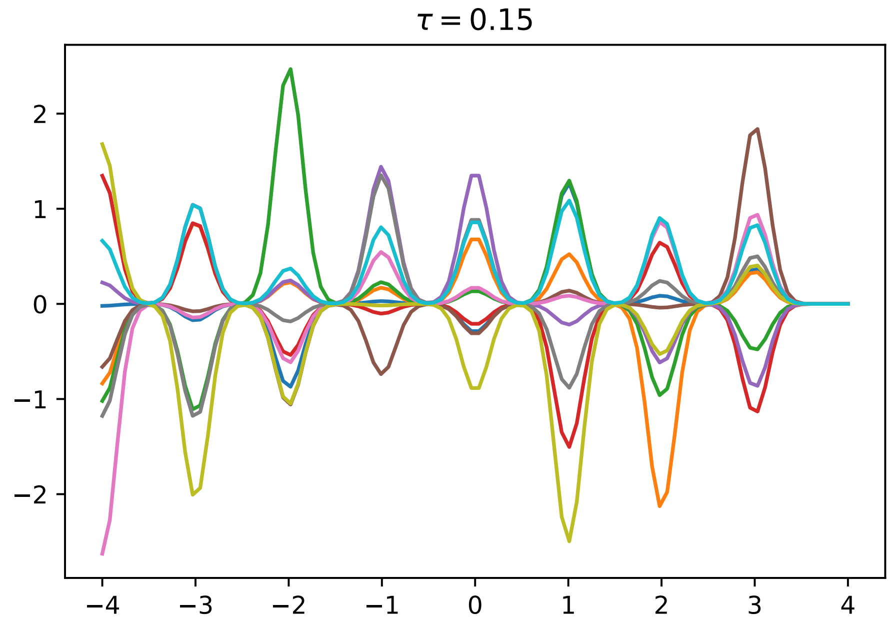
$$C = \frac{1}{(2\pi)^{\frac{2m+1}{2}}|\boldsymbol{K}|^{\frac{1}{2}}}$$

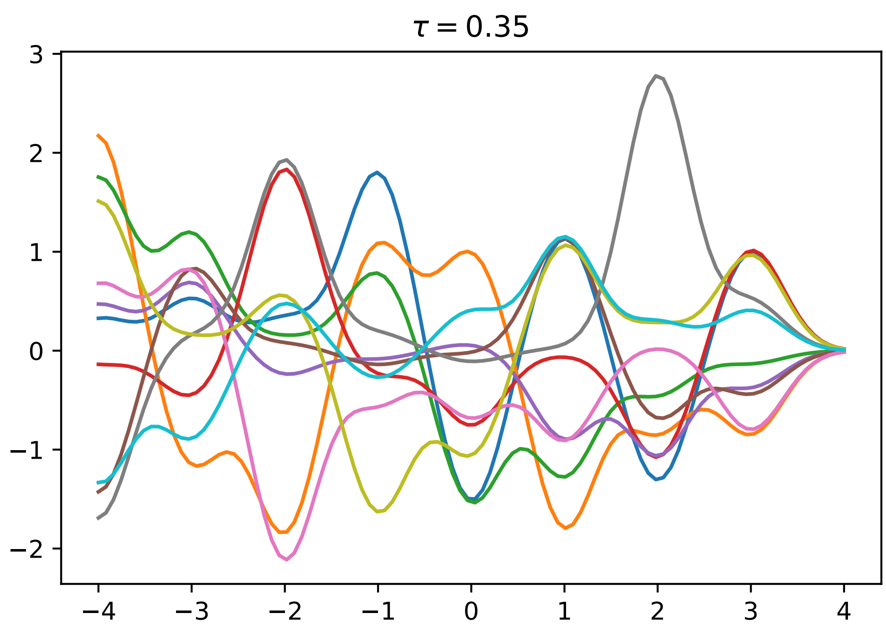
which is taken from the $\mathcal{N}(\boldsymbol{f}; \boldsymbol{0}, \boldsymbol{K})$ normalization constant

3. (a) kernel matrix: $\boldsymbol{K} = [k_{ij}] = e^{-\frac{(i-j)^2}{2\tau^2}}$

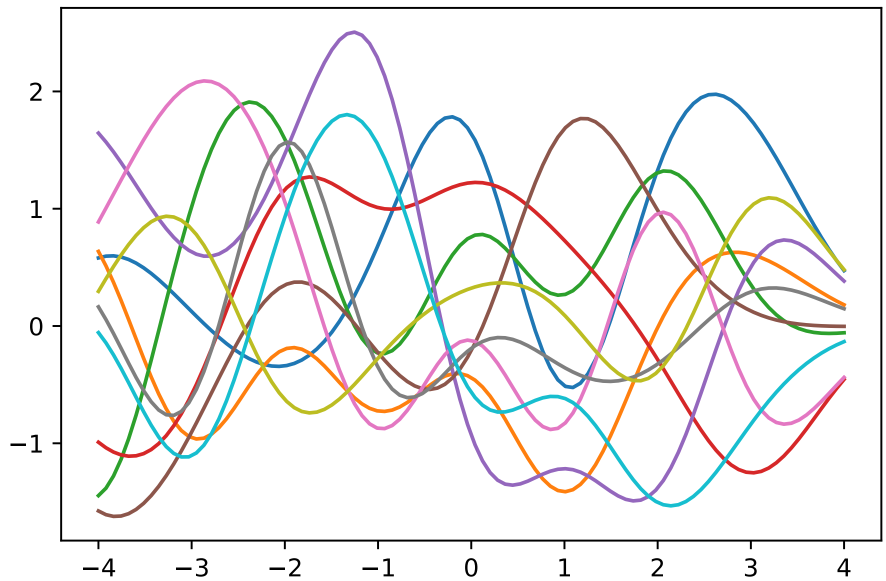


(b) functions become smoother as τ increases

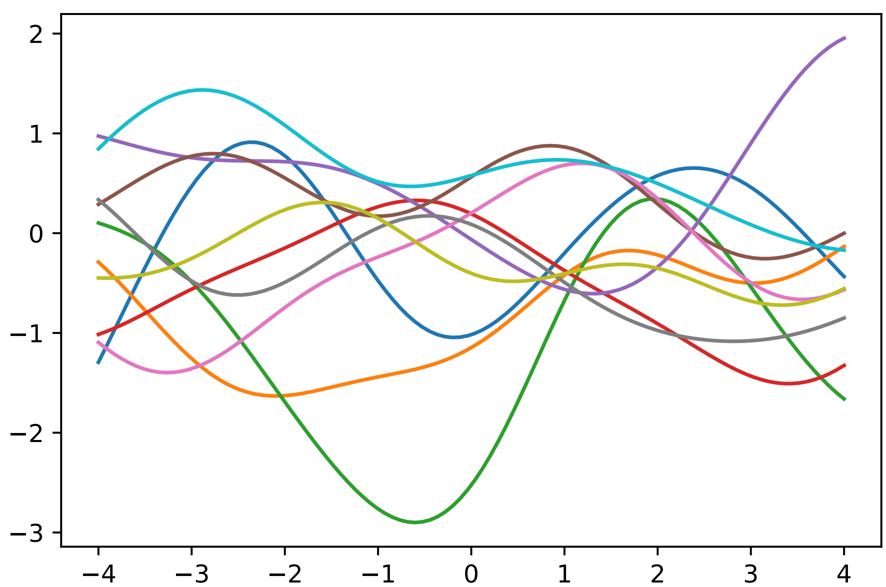


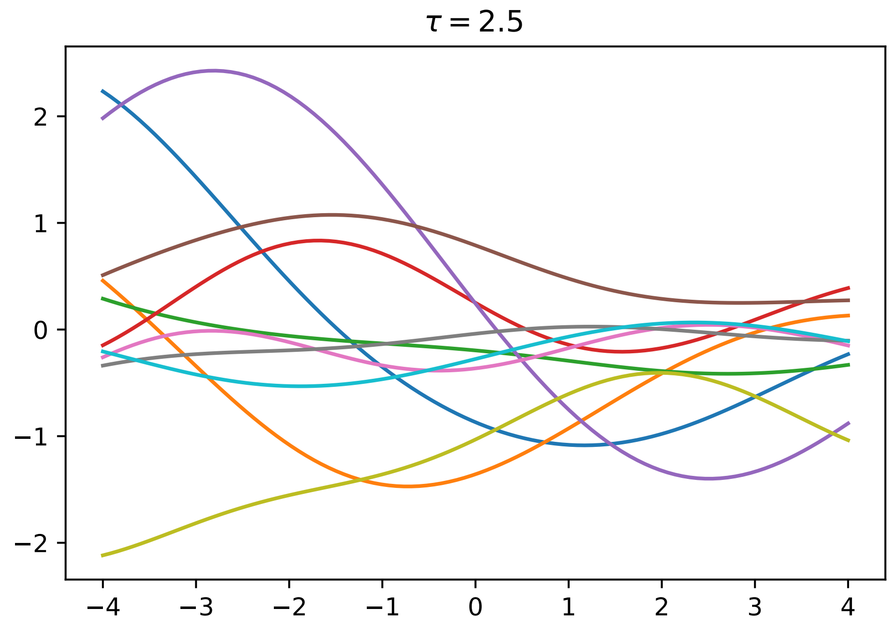


$\tau = 0.7$

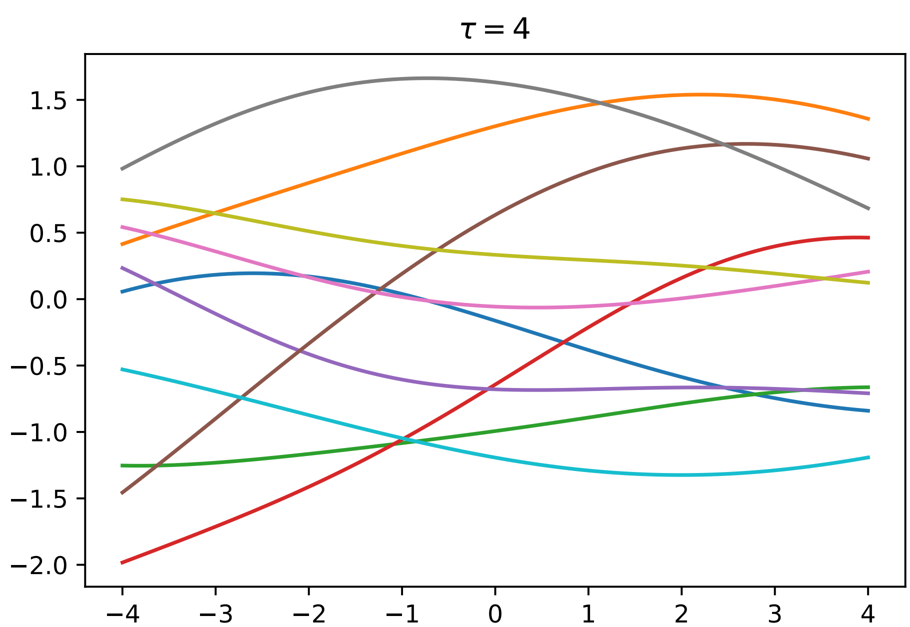


$\tau = 1.5$

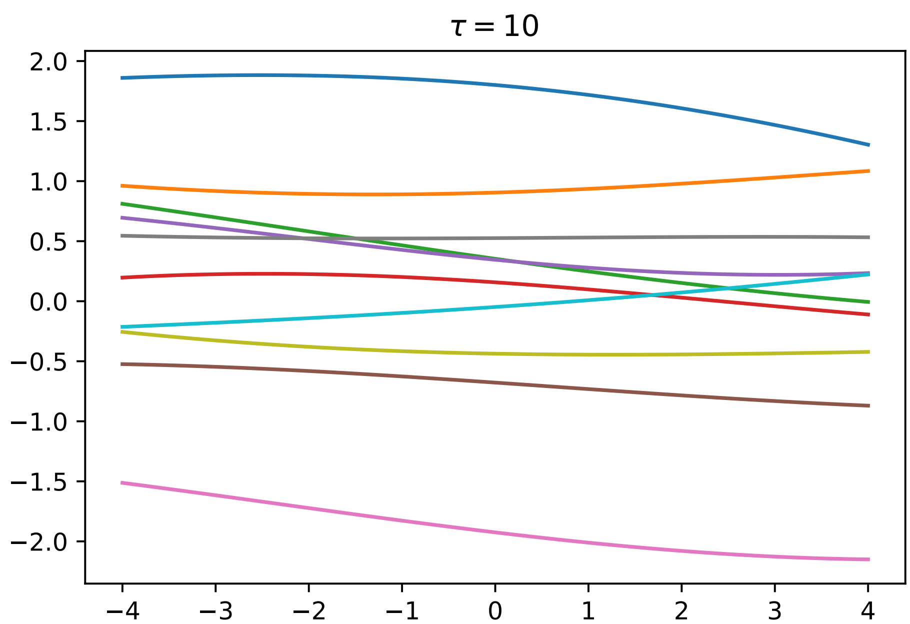




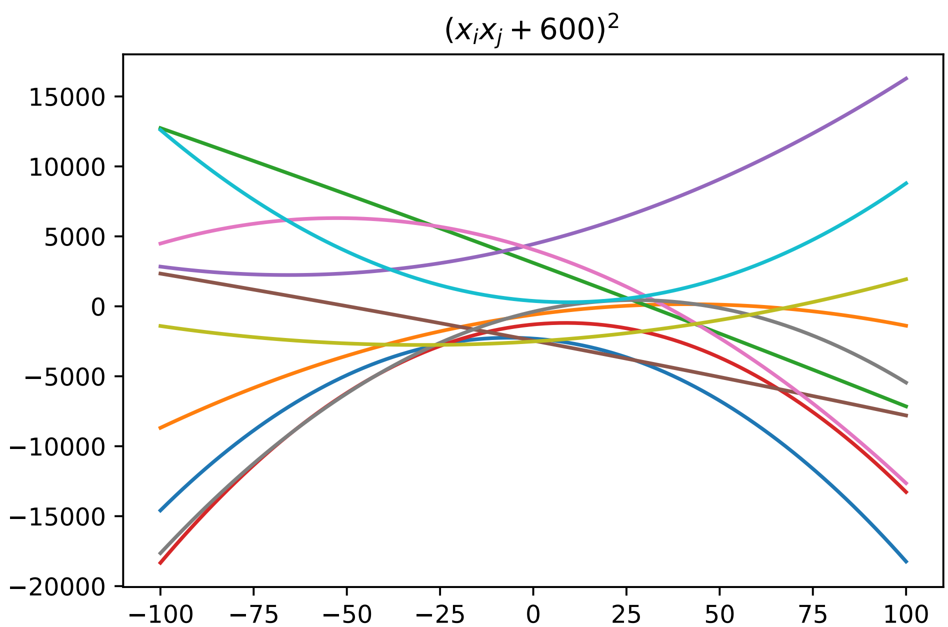
$\tau = 2.5$
2
1
0
−1
−2
−4
−3
−2
−1
0
1
2
3
4

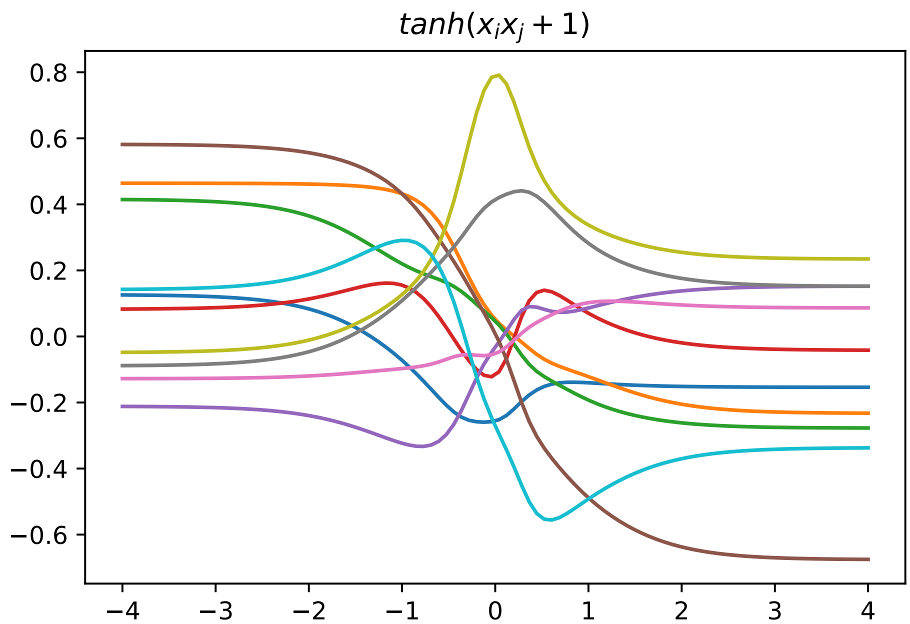


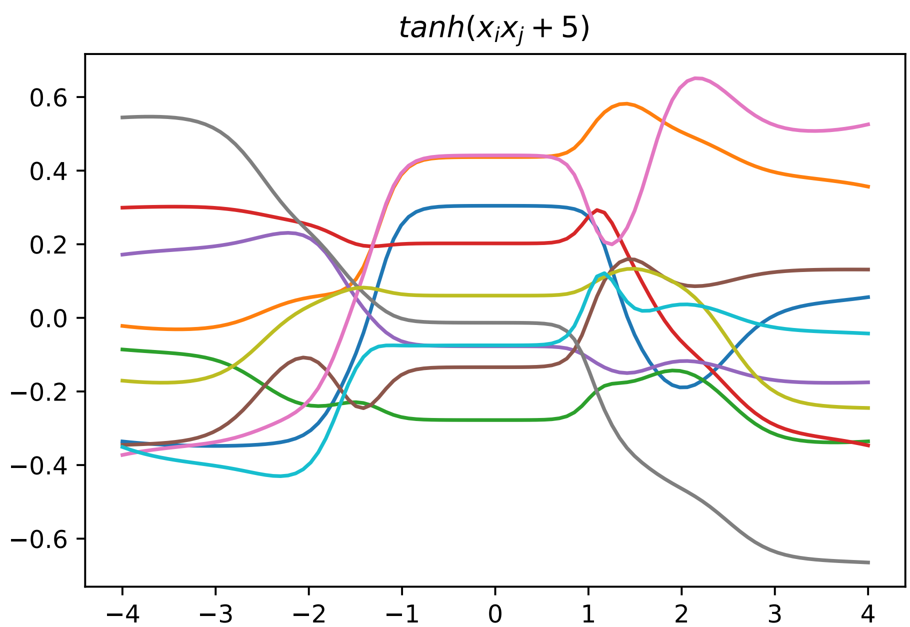
$\tau = 4$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−4
−3
−2
−1
0
1
2
3
4

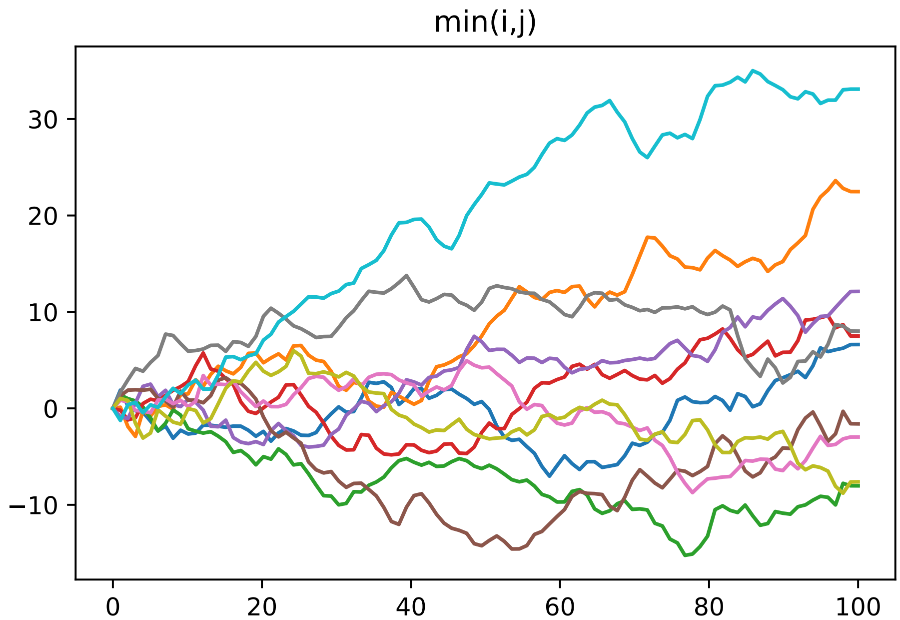

$\tau = 10$
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−4
−3
−2
−1
0
1
2
3
4

(c)

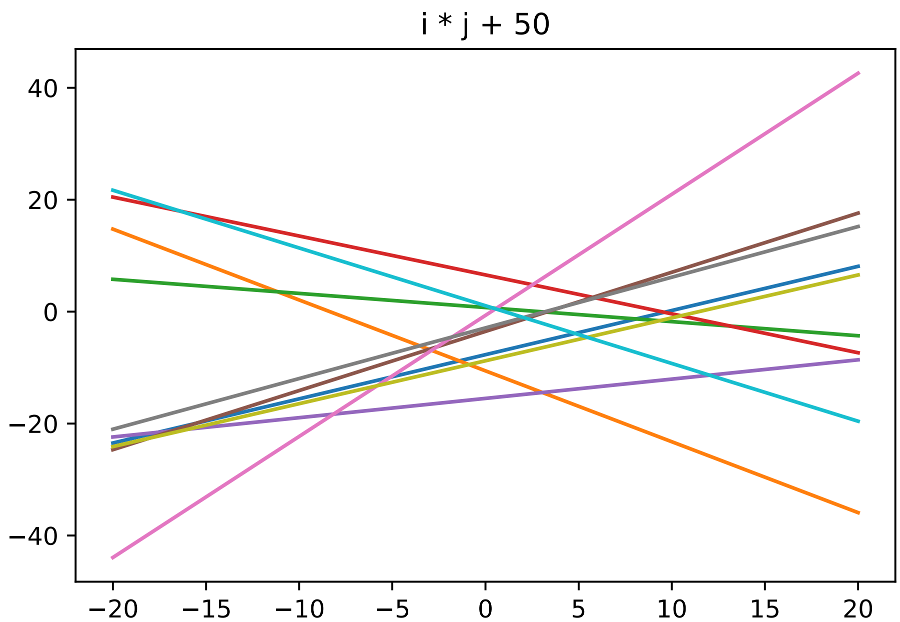

$(x_i x_j + 600)^2$
15000
10000
5000
0
−5000
−10000
−15000
−20000
−100
−75
−50
−25
0
25
50
75
100


$tanh(x_i x_j + 1)$
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−4
−3
−2
−1
0
1
2
3
4


$tanh(x_i x_j + 5)$
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−4
−3
−2
−1
0
1
2
3
4



min(i,j)
30
20
10
0
−10
0
20
40
60
80
100



i * j + 50
40
20
0
−20
−40
−20
−15
−10
−5
0
5
10
15
20