Wasserstein Consensus ADMM

Iman Nodozi (inodozi@ucsc.edu), Abhishek Halder (ahalder@ucsc.edu)

University of California, Santa Cruz

Main Idea

$$\underset{\mu\in\mathcal{P}_2(\mathbb{R}^d)}{\arg\inf}\; F_1(\mu)+F_2(\mu)+\ldots+F_n(\mu)$$

re-write

$$\underset{(\mu_1,\ldots,\mu_n,\zeta)\in\mathcal{P}_2^{n+1}(\mathbb{R}^d)}{\arg\inf}\; F_1(\mu_1)+F_2(\mu_2)+\ldots+F_n(\mu_n)$$

subject to $\mu_i=\zeta$ for all $i\in[n]$

Define Wasserstein augmented Lagrangian:

$$L_\alpha(\mu_1,\ldots,\mu_n,\zeta,\nu_1,\ldots,\nu_n):=\sum_{i=1}^{n}\left\{F_i(\mu_i)+\frac{\alpha}{2}W^2(\mu_i,\zeta)+\int_{\mathbb{R}^d}\nu_i(\boldsymbol{\theta})(\mathrm{d}\mu_i-\mathrm{d}\zeta)\right\}$$

regularization > 0 — Lagrange multipliers

$$\mu_i^{k+1}=\underset{\mu_i\in\mathcal{P}_2(\mathbb{R}^d)}{\arg\inf}\; L_\alpha\left(\mu_1,\ldots,\mu_n,\zeta^k,\nu_1^k,\ldots,\nu_n^k\right)$$

$$\zeta^{k+1}=\underset{\zeta\in\mathcal{P}_2(\mathbb{R}^d)}{\arg\inf}\; L_\alpha\left(\mu_1^{k+1},\ldots,\mu_n^{k+1},\zeta,\nu_1^k,\ldots,\nu_n^k\right)$$

$$\nu_i^{k+1}=\nu_i^k+\alpha\left(\mu_i^{k+1}-\zeta^{k+1}\right)$$

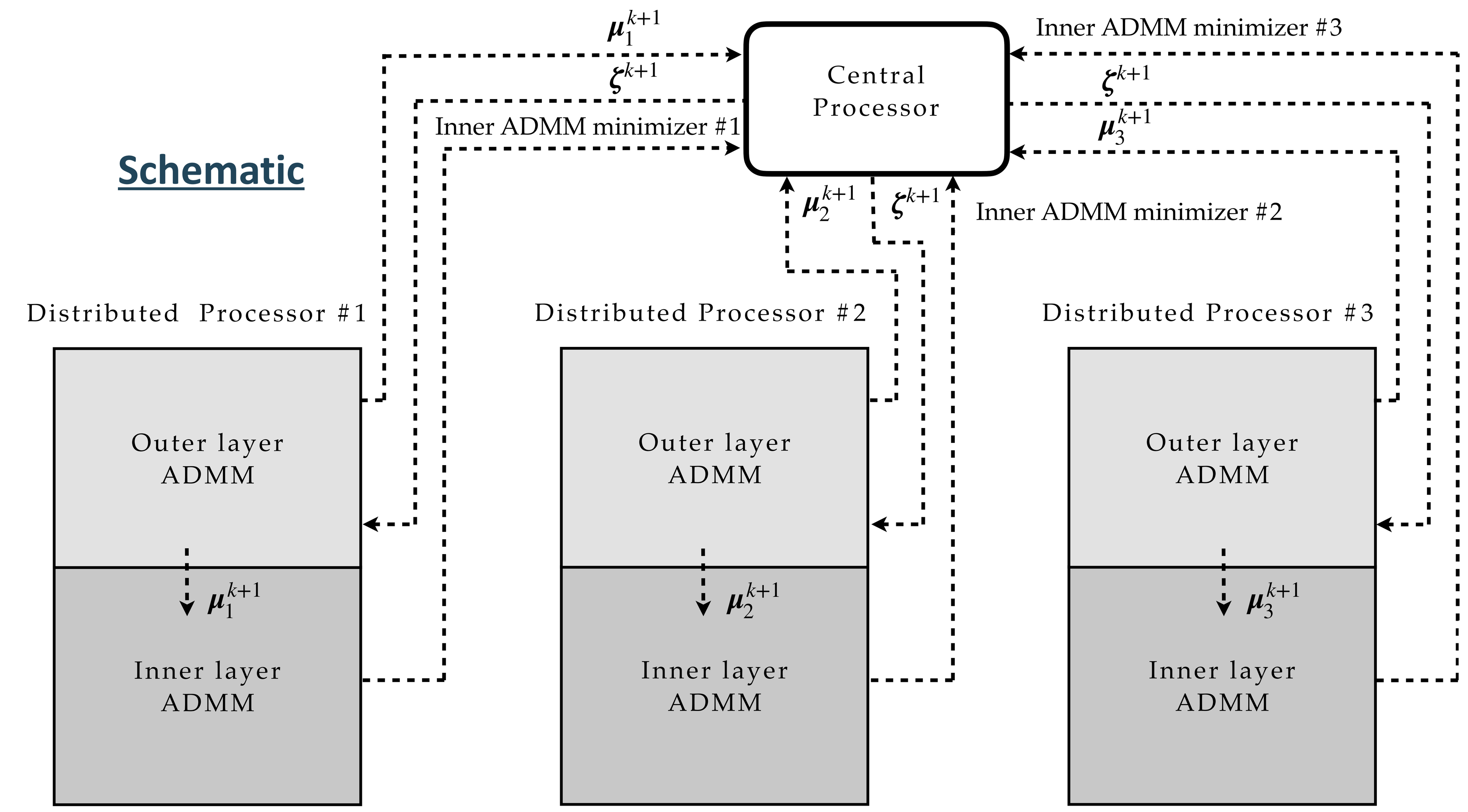
and simplify the recursions to

$$\mu_i^{k+1}=\mathrm{prox}^{W}_{\frac{1}{\alpha}\left(F_i(\cdot)+\int\nu_i^k\,\mathrm{d}(\cdot)\right)}\left(\zeta^k\right)$$

$$\zeta^{k+1}=\underset{\zeta\in\mathcal{P}_2(\mathbb{R}^d)}{\arg\inf}\left\{\left(\sum_{i=1}^{n}W^2\left(\mu_i^{k+1},\zeta\right)\right)-\frac{2}{\alpha}\int_{\mathbb{R}^d}\nu_{\mathrm{sum}}^k(\boldsymbol{\theta})\mathrm{d}\zeta\right\}$$

$$\nu_i^{k+1}=\nu_i^k+\alpha\left(\mu_i^{k+1}-\zeta^{k+1}\right)$$

Schematic



Examples:

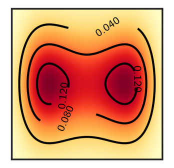
$\Phi_i(\cdot)=F_i(\cdot)+\int\nu_i^k\mathrm{d}(\cdot)$	PDE	Name
$\int_{\mathbb{R}^d}\left(V(\boldsymbol{\theta})+\nu_i^k(\boldsymbol{\theta})\right)\mathrm{d}\mu_i(\boldsymbol{\theta})$	$\frac{\partial\widetilde{\mu}_i}{\partial t}=\nabla\cdot\left(\widetilde{\mu}_i\left(\nabla V+\nabla\nu_i^k\right)\right)$	Liouville equation
$\int_{\mathbb{R}^d}\left(\nu_i^k(\boldsymbol{\theta})+\beta^{-1}\log\mu_i(\boldsymbol{\theta})\right)\mathrm{d}\mu_i(\boldsymbol{\theta})$	$\frac{\partial\widetilde{\mu}_i}{\partial t}=\nabla\cdot\left(\widetilde{\mu}_i\nabla\nu_i^k\right)+\beta^{-1}\Delta\widetilde{\mu}_i$	Fokker-Planck equation
$\int_{\mathbb{R}^d}\nu_i^k(\boldsymbol{\theta})\mathrm{d}\mu_i(\boldsymbol{\theta})+\int_{\mathbb{R}^{2d}}U(\boldsymbol{\theta},\boldsymbol{\sigma})\mathrm{d}\mu_i(\boldsymbol{\theta})\mathrm{d}\mu_i(\boldsymbol{\sigma})$	$\frac{\partial\widetilde{\mu}_i}{\partial t}=\nabla\cdot\left(\widetilde{\mu}_i\left(\nabla\nu_i^k+\nabla\left(U\circledast\widetilde{\mu}_i\right)\right)\right)$	Propagation of chaos equation
$\int_{\mathbb{R}^d}\left(\nu_i^k(\boldsymbol{\theta})+\frac{\beta^{-1}}{m-1}\mathbf{1}^\top\mu_i^m\right)\mathrm{d}\mu_i(\boldsymbol{\theta}),m>1$	$\frac{\partial\widetilde{\mu}_i}{\partial t}=\nabla\cdot\left(\widetilde{\mu}_i\nabla\nu_i^k\right)+\beta^{-1}\Delta\widetilde{\mu}_i^m$	Porous medium equation

Linear Fokker-Planck-Kolmogorov PDE

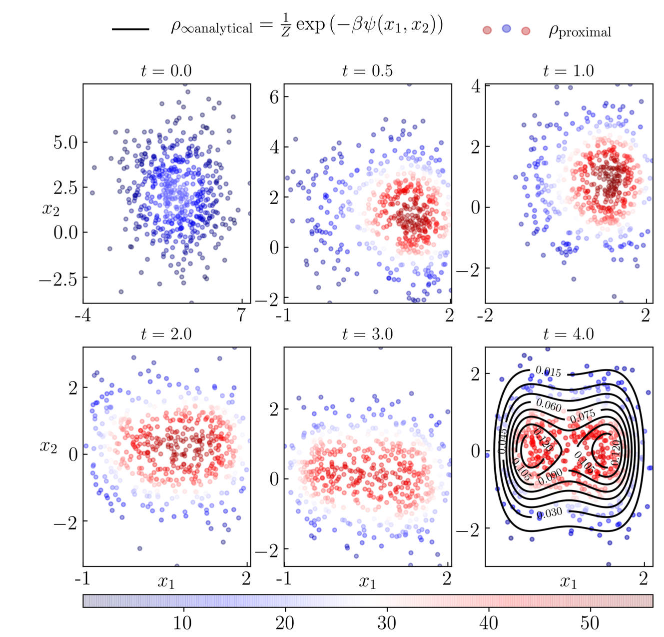
$$\frac{\partial\mu}{\partial t}=\nabla\cdot(\mu\nabla V)+\beta^{-1}\Delta\mu$$

$$V(x_1,x_2)=\tfrac{1}{4}\left(1+x_1^4\right)+\tfrac{1}{2}\left(x_2^2-x_1^2\right)$$

$$\mu_\infty\propto\exp(-\beta V(x_1,x_2))\mathrm{d}x_1\mathrm{d}x_2$$



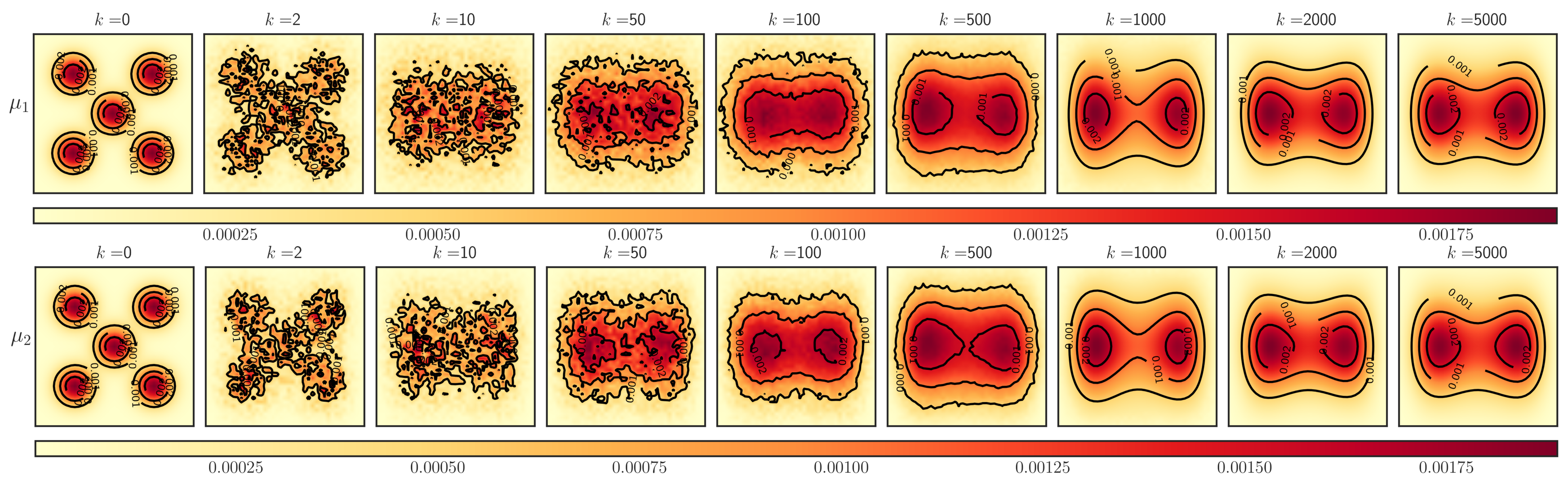
Centralized computation:



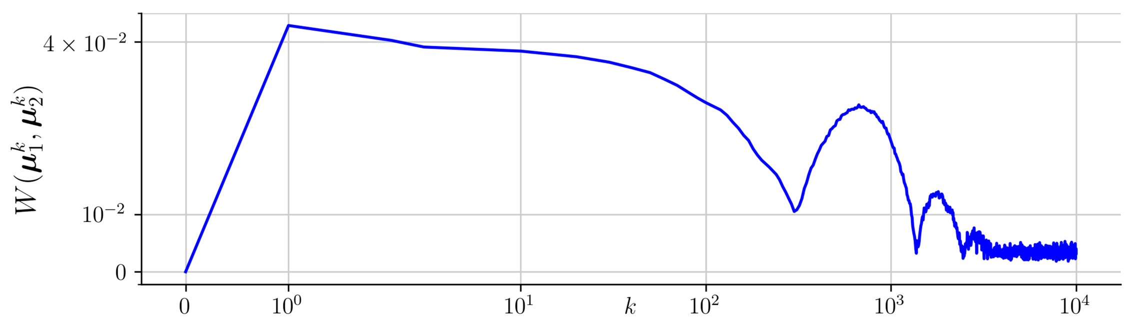
Runtime 99.89 s on Macbook Air 1.1 GHz intel i5 8GB RAM

Distributed computation:

$F_1(\boldsymbol{\mu})=\langle\boldsymbol{V}_k,\boldsymbol{\mu}\rangle$ $F_2(\boldsymbol{\mu})=\langle\beta^{-1}\log\boldsymbol{\mu},\boldsymbol{\mu}\rangle$



Wasserstein distance



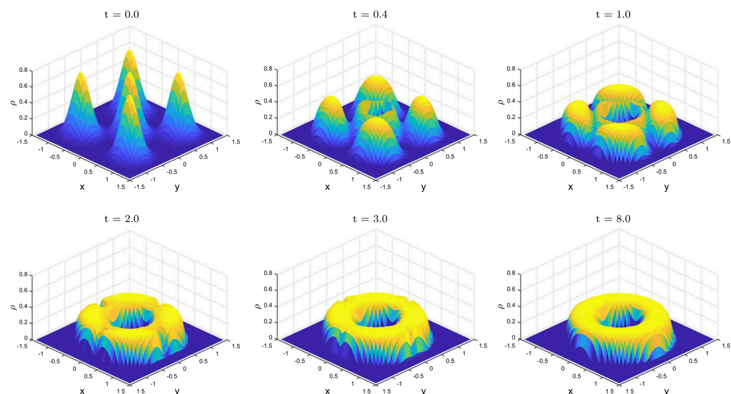
Centralized computation:

Aggregation-drift-diffusion nonlinear PDE

$$\frac{\partial\mu}{\partial t}=\underbrace{\nabla\cdot(\mu\nabla(U*\mu))}_{i=1}+\underbrace{\nabla\cdot(\mu\nabla V)+\beta^{-1}\Delta\mu^2}_{i=2}$$

$$U(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{x}\|_2^2-\ln\|\boldsymbol{x}\|_2$$

$$V(\boldsymbol{x})=-\frac{1}{4}\ln\|\boldsymbol{x}\|_2$$

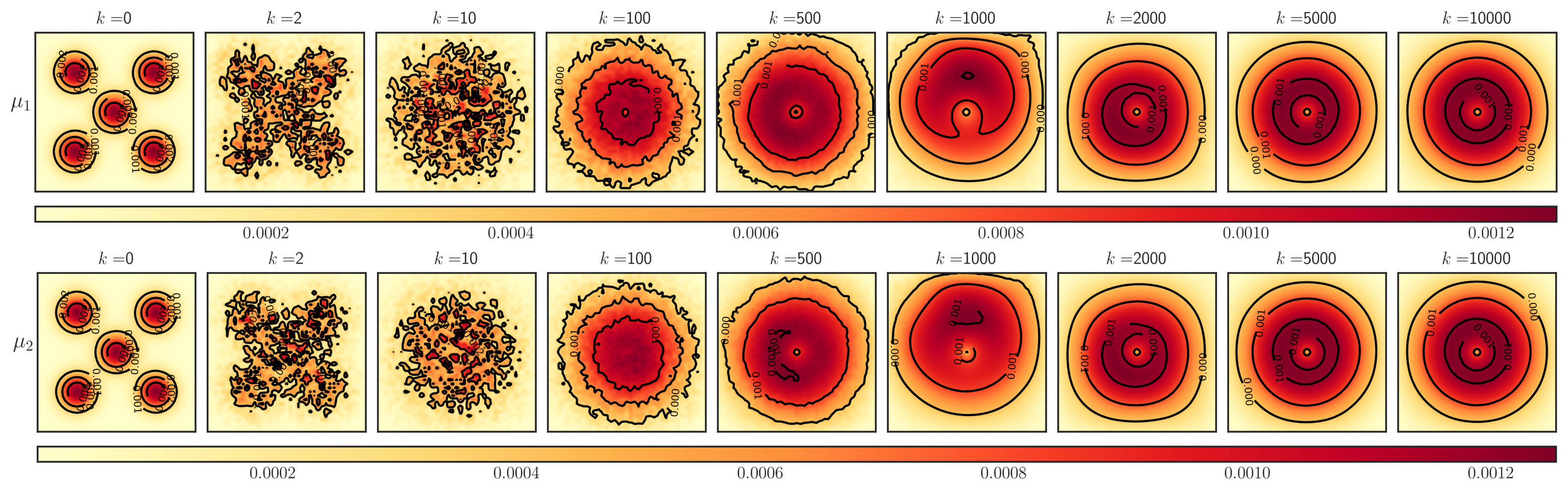


$\lim_{\beta^{-1}\downarrow 0}\mu_\infty=\mathrm{Unif}(\mathscr{A})$

Annulus with inner radius 1/2 and outer radius $\sqrt{5}/2$

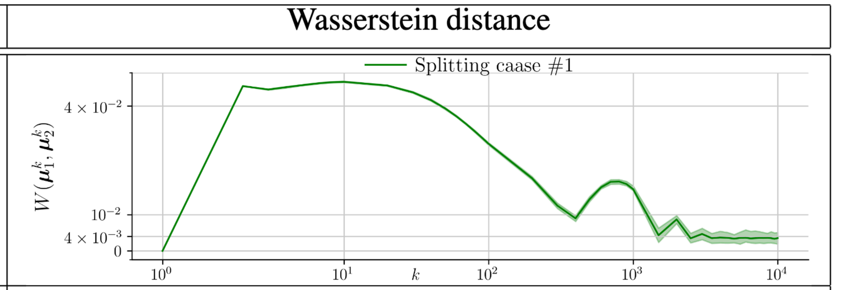
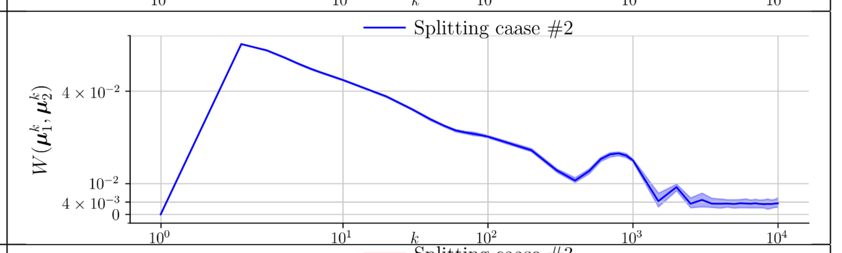
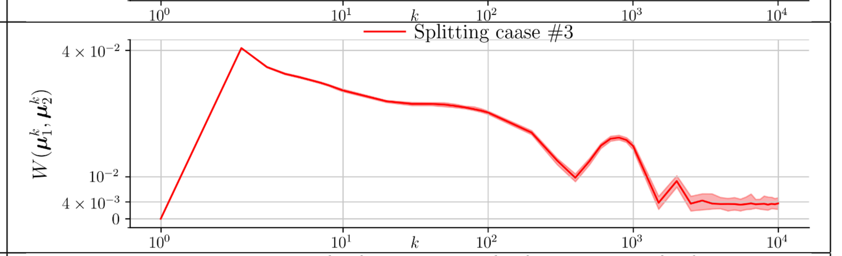
Distributed computation:

$F_1(\boldsymbol{\mu})=\langle\boldsymbol{U}_k\boldsymbol{\mu},\boldsymbol{\mu}\rangle$ $F_2(\boldsymbol{\mu})=\langle\boldsymbol{V}_k+\beta^{-1}\log\boldsymbol{\mu},\boldsymbol{\mu}\rangle$



Wasserstein distance

100 run for statistics each of the 4 ways of splitting: (2^n-n-1 ways in general)

Splitting case	Functionals	Wasserstein distance
#1	$F_1(\boldsymbol{\mu})=\left\langle\boldsymbol{V}_k+\beta^{-1}\boldsymbol{\mu},\boldsymbol{\mu}\right\rangle$, $F_2(\boldsymbol{\mu})=\left\langle\boldsymbol{U}_k\boldsymbol{\mu}^k,\boldsymbol{\mu}\right\rangle$ av. runtime = 294.06 s	
#2	$F_1(\boldsymbol{\mu})=\langle\boldsymbol{U}_k\boldsymbol{\mu}^k+\beta^{-1}\boldsymbol{\mu},\boldsymbol{\mu}\rangle$, $F_2(\boldsymbol{\mu})=\langle\boldsymbol{V}_k,\boldsymbol{\mu}\rangle$ av. runtime = 285.32 s	
#3	$F_1(\boldsymbol{\mu})=\langle\boldsymbol{U}_k\boldsymbol{\mu}^k+\boldsymbol{V}_k,\boldsymbol{\mu}\rangle$, $F_2(\boldsymbol{\mu})=\left\langle\beta^{-1}\boldsymbol{\mu},\boldsymbol{\mu}\right\rangle$ av. runtime = 289.87 s	
#4	$F_1(\boldsymbol{\mu})=\langle\boldsymbol{V}_k,\boldsymbol{\mu}\rangle$, $F_2(\boldsymbol{\mu})=\left\langle\boldsymbol{U}_k\boldsymbol{\mu}^k\right\rangle$, $F_3(\boldsymbol{\mu})=\left\langle\beta^{-1}\boldsymbol{\mu},\boldsymbol{\mu}\right\rangle$ av. runtime = 108.99 s