Analysis of Momentum Methods

Nikola B. Kovachki[1] Andrew M. Stuart[1]

[1]Computational and Mathematical Sciences
California Institute of Technology

SIAM DS19
May 19-23th, 2019

Table of Contents

Table of Contents

Optimization for ML

- **Solve**:

$$\arg\min_{u \in \mathbb{R}^d} \; \Phi(u)$$

$\Phi(\cdot)$ non-convex objective.

Optimization for ML

- **Solve**:

$$\arg\min_{u \in \mathbb{R}^d} \; \Phi(u)$$

  $\Phi(\cdot)$ non-convex objective.

- **Gradient flow**:

$$\frac{du}{dt} = -\nabla\Phi(u)$$

  will decrease Φ along trajectories.

Optimization for ML

- **Solve**:

$$\arg\min_{u \in \mathbb{R}^d} \; \Phi(u)$$

  $\Phi(\cdot)$ non-convex objective.

- **Gradient flow**:

$$\frac{du}{dt} = -\nabla \Phi(u)$$

  will decrease Φ along trajectories.

- **Gradient decent**:

$$\mathbf{u}_{n+1} = \mathbf{u}_n - h \nabla \Phi(\mathbf{u}_n)$$

  $0 < h < 1$ "learning rate."

Momentum Methods

- **Heavy Ball** (Polyak 1964):

$$\mathbf{v}_{n+1} = \lambda_n \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n)$$
$$\mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{v}_{n+1}$$

Momentum Methods

- **Heavy Ball** (Polyak 1964):

$$\mathbf{v}_{n+1} = \lambda_n \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n)$$
$$\mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{v}_{n+1}$$

- **Nesterov's Accelerated Gradients** (Nesterov 1983):

$$\mathbf{v}_{n+1} = \lambda_n \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + \lambda_n \mathbf{v}_n)$$
$$\mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{v}_{n+1}$$

$\lambda_n = n/(n+3)$.

Momentum Methods

- **Heavy Ball** (Polyak 1964):

$$\mathbf{v}_{n+1} = \lambda_n \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n)$$
$$\mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{v}_{n+1}$$

- **Nesterov's Accelerated Gradients** (Nesterov 1983):

$$\mathbf{v}_{n+1} = \lambda_n \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + \lambda_n \mathbf{v}_n)$$
$$\mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{v}_{n+1}$$

  $\lambda_n = n/(n+3)$.

- **Link**:

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda_n(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a_n(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

  $a_n = 0$ HB, $a_n = \lambda_n$ NAG.

Physical Analogy

- **Heavy Ball** (Qian 1999):

$$m\frac{d^2u}{dt^2} + \gamma(t)\frac{du}{dt} + \nabla\Phi(u) = 0$$

Physical Analogy

- **Heavy Ball** (Qian 1999):

$$m\frac{d^2u}{dt^2} + \gamma(t)\frac{du}{dt} + \nabla\Phi(u) = 0$$

- **Nesterov's Accelerated Gradients** (Su et al. 2014):

$$\frac{d^2u}{dt^2} + \frac{3}{t}\frac{du}{dt} + \nabla\Phi(u) = 0$$

In Deep Learning

- **Use Case**:

$$\lambda_n = \lambda \in (0, 1)$$

In Deep Learning

- **Use Case**:
$$\lambda_n = \lambda \in (0, 1)$$
- Effect on continuum dynamic?

Table of Contents

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$
$$\frac{du}{dt} - \lambda\frac{du}{dt} = -\nabla\Phi(u)$$

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$
$$\frac{du}{dt} - \lambda\frac{du}{dt} = -\nabla\Phi(u)$$

- **Discretize**:

$$\frac{u(t+h)-u(t)}{h} - \lambda\frac{u(t)-u(t-h)}{h} \approx -\nabla\Phi(u(t))$$

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$
$$\frac{du}{dt} - \lambda\frac{du}{dt} = -\nabla\Phi(u)$$

- **Discretize**:

$$\frac{u(t+h)-u(t)}{h} - \lambda\frac{u(t)-u(t-h)}{h} \approx -\nabla\Phi(u(t))$$
$$-\nabla\Phi(u(t)) \approx -\nabla\Phi\left(u(t) + ha\frac{u(t)-u(t-h)}{h}\right)$$

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$
$$\frac{du}{dt} - \lambda\frac{du}{dt} = -\nabla\Phi(u)$$

- **Discretize**:

$$\frac{u(t+h)-u(t)}{h} - \lambda\frac{u(t)-u(t-h)}{h} \approx -\nabla\Phi(u(t))$$
$$-\nabla\Phi(u(t)) \approx -\nabla\Phi\left(u(t) + ha\frac{u(t)-u(t-h)}{h}\right)$$

$$u(t+h) \approx u(t)+\lambda(u(t)-u(t-h))-h\nabla\Phi(u(t)+a(u(t)-u(t-h)))$$

Intuition

- **Re-scaled Gradient Flow**:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$
$$\frac{du}{dt} - \lambda\frac{du}{dt} = -\nabla\Phi(u)$$

- **Discretize**:

$$\frac{u(t+h)-u(t)}{h} - \lambda\frac{u(t)-u(t-h)}{h} \approx -\nabla\Phi(u(t))$$
$$-\nabla\Phi(u(t)) \approx -\nabla\Phi\left(u(t) + ha\frac{u(t)-u(t-h)}{h}\right)$$

$$u(t+h) \approx u(t)+\lambda(u(t)-u(t-h))-h\nabla\Phi(u(t)+a(u(t)-u(t-h)))$$

- **Recall**:

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1})) \quad (1)$$

Convergence

Theorem

Suppose $\Phi \in C_b^3(\mathbb{R}^d;\mathbb{R})$ *and let* $u \in C^3([0,\infty);\mathbb{R}^d)$ *be the solution to*

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$

with $\lambda \in (0,1)$. *For* $n = 0,1,2,\ldots$ *let* $\mathbf{u}_n$ *be the sequence given by* (1) *and define* $u_n = u(nh)$. *The for any* $T \geq 0$, *there is a constant* $C = C(T) > 0$ *such that*

$$\sup_{0 \leq nh \leq T} |u_n - \mathbf{u}_n| \leq Ch$$

Numerical Illustration

- $\Phi(u) = \frac{1}{2}u^2$, $u(0) = \mathbf{u}_0 = 1$.

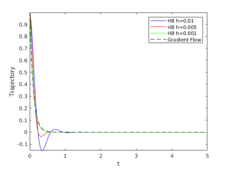

Figure: $\lambda = 0.9$

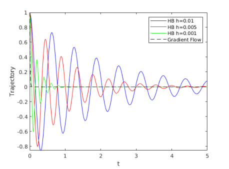

Figure: $\lambda = 0.99$

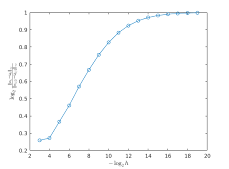

Figure: OOC

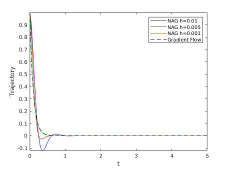

Figure: $\lambda = 0.9$

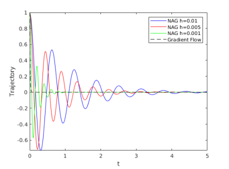

Figure: $\lambda = 0.99$

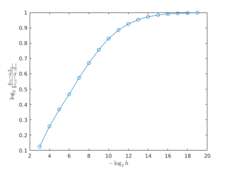

Figure: OOC

Table of Contents

Intuition

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

Intuition

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Add and Subtract**: $\mathbf{u}_n - \mathbf{u}_{n-1}$

$$h\frac{\mathbf{u}_{n+1} - 2\mathbf{u}_n + \mathbf{u}_{n-1}}{h^2} + (1-\lambda)\frac{\mathbf{u}_n - \mathbf{u}_{n-1}}{h} = -\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

Intuition

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Add and Subtract**: $\mathbf{u}_n - \mathbf{u}_{n-1}$

$$h\frac{\mathbf{u}_{n+1} - 2\mathbf{u}_n + \mathbf{u}_{n-1}}{h^2} + (1-\lambda)\frac{\mathbf{u}_n - \mathbf{u}_{n-1}}{h} = -\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Hamiltonian dynamic**:

$$h\frac{d^2u}{dt^2} + (1-\lambda)\frac{du}{dt} + \nabla\Phi(u) = 0$$

Convergence

Theorem

Suppose $\Phi \in C_b^3(\mathbb{R}^d;\mathbb{R})$*, fix* $\lambda \in (0,1)$ *and suppose* $a \geq 0$ *is chosen so that* $\alpha = \frac{1}{2}(1+\lambda-2a(1-\lambda))$ *is strictly positive. Let* $u \in C^4([0,\infty),\mathbb{R}^d)$ *be the solution to*

$$h\alpha \frac{d^2u}{dt^2} + (1-\lambda)\frac{du}{dt} = -\nabla\Phi(u)$$

with h *small enough. For* $n = 0,1,2,\dots$ *let* $\mathbf{u}_n$ *be the sequence given by* (1) *and define* $u_n = u(nh)$*. Then for any* $T \geq 0$*, there is a constant* $C = C(T) > 0$ *such that*

$$\sup_{0 \leq nh \leq T} |u_n - \mathbf{u}_n| \leq Ch$$

First Order Approximation

Lemma

Suppose $\Phi \in C_b^3(\mathbb{R}^d;\mathbb{R})$*, fix* $\lambda \in (0,1)$ *and* $a, \alpha \geq 0$*. Let* $u \in C^4([0,\infty),\mathbb{R}^d)$ *be the solution to*

$$h\alpha \frac{d^2 u}{dt^2} + (1-\lambda)\frac{du}{dt} = -\nabla\Phi(u).$$

Then, if h *is small enough, there are constants* $C^{(1)}, C_1^{(2)}, C_2^{(2)}, C_1^{(3)}, C_2^{(3)} > 0$ *independent of* h *such that for any* $t \in [0,\infty)$

$$|\dot{u}(t)| \leq C^{(1)}$$

$$|\ddot{u}(t)| \leq \frac{C_1^{(2)}}{h} exp\left(-\frac{1-\lambda}{2h\alpha}t\right) + C_2^{(2)}$$

$$|\dddot{u}(t)| \leq \frac{C_1^{(3)}}{h^2} exp\left(-\frac{1-\lambda}{2h\alpha}t\right) + C_2^{(3)}$$

Numerical Illustration

- $\Phi(u) = \frac{1}{2}u^2,\ u(0) = \mathbf{u}_0 = 1,\ \dot{u}(0) = -\frac{1-2\alpha}{2\alpha-\lambda+1}\nabla\Phi(\mathbf{u}_0).$

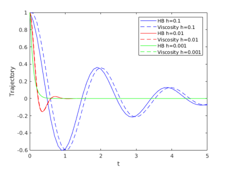

Figure: $\lambda = 0.9$

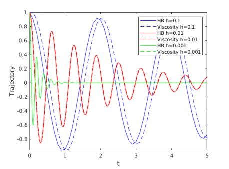

Figure: $\lambda = 0.9$

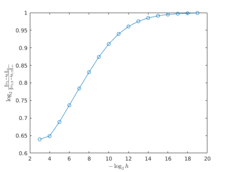

Figure: OOC

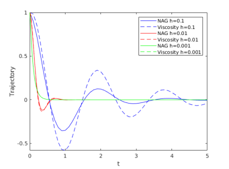

Figure: $\lambda = 0.9$

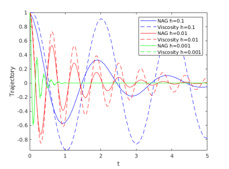

Figure: $\lambda = 0.99$

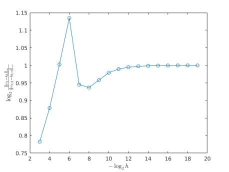

Figure: OOC

Higher Order Equations

- **ODE(s)**:

$$\sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k} = -\nabla\Phi(u)$$

Higher Order Equations

- **ODE(s)**:

$$\sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k} = -\nabla\Phi(u)$$

- **Taylor Expansion**:

$$u_{n+1} = u_n + \lambda(u_n - u_{n-1}) - h\nabla\Phi(u_n) + Ch^{p+1}$$

where $u_n = u(nh)$

Higher Order Equations

- **ODE(s)**:

$$\sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!} \frac{d^k u}{dt^k} = -\nabla\Phi(u)$$

- **Taylor Expansion**:

$$u_{n+1} = u_n + \lambda(u_n - u_{n-1}) - h\nabla\Phi(u_n) + Ch^{p+1}$$

where $u_n = u(nh)$

- **Order 1**:

$$C \propto \frac{1}{h^{p-1}}$$

Higher Order Equations

- **Operators**: $T_p : C^\infty([0,\infty),\mathbb{R}^d) \to C^\infty([0,\infty),\mathbb{R}^d)$

$$T_p u = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k}, \quad \forall u \in C^\infty([0,\infty),\mathbb{R}^d)$$

Higher Order Equations

- **Operators**: $T_p : C^\infty([0,\infty),\mathbb{R}^d) \to C^\infty([0,\infty),\mathbb{R}^d)$

$$T_p u = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k}, \quad \forall u \in C^\infty([0,\infty),\mathbb{R}^d)$$

- **Fourier Transform**:

$$\mathcal{F}(T_p u)(\omega) = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)(i\omega)^k}{k!}\mathcal{F}(u)(\omega)$$

Higher Order Equations

- **Operators**: $T_p : C^\infty([0,\infty),\mathbb{R}^d) \to C^\infty([0,\infty),\mathbb{R}^d)$

$$T_p u = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k}, \quad \forall u \in C^\infty([0,\infty),\mathbb{R}^d)$$

- **Fourier Transform**:

$$\mathcal{F}(T_p u)(\omega) = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)(i\omega)^k}{k!}\mathcal{F}(u)(\omega)$$

$$\mathcal{F}(Tu)(\omega) = \frac{1}{h}(e^{ih\omega} + \lambda e^{-ih\omega} - \lambda - 1)\mathcal{F}(u)(\omega) \qquad (p \to \infty)$$

Higher Order Equations

- **Operators**: $T_p : C^\infty([0,\infty),\mathbb{R}^d) \to C^\infty([0,\infty),\mathbb{R}^d)$

$$T_p u = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k}, \quad \forall u \in C^\infty([0,\infty),\mathbb{R}^d)$$

- **Fourier Transform**:

$$\mathcal{F}(T_p u)(\omega) = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)(i\omega)^k}{k!}\mathcal{F}(u)(\omega)$$

$$\mathcal{F}(Tu)(\omega) = \frac{1}{h}(e^{ih\omega} + \lambda e^{-ih\omega} - \lambda - 1)\mathcal{F}(u)(\omega) \qquad (p \to \infty)$$

$$(Tu)(t) = \frac{1}{h}\mathcal{F}^{-1}(e^{ih\omega} + \lambda e^{-ih\omega} - \lambda - 1)(t) * u(t)$$

Higher Order Equations

- **Operators**: $T_p : C^\infty([0,\infty),\mathbb{R}^d) \to C^\infty([0,\infty),\mathbb{R}^d)$

$$T_p u = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)}{k!}\frac{d^k u}{dt^k}, \quad \forall u \in C^\infty([0,\infty),\mathbb{R}^d)$$

- **Fourier Transform**:

$$\mathcal{F}(T_p u)(\omega) = \sum_{k=1}^{p} \frac{h^{k-1}(1+(-1)^k\lambda)(i\omega)^k}{k!}\mathcal{F}(u)(\omega)$$

$$\mathcal{F}(Tu)(\omega) = \frac{1}{h}(e^{ih\omega} + \lambda e^{-ih\omega} - \lambda - 1)\mathcal{F}(u)(\omega) \qquad (p \to \infty)$$

$$(Tu)(t) = \frac{1}{h}\mathcal{F}^{-1}(e^{ih\omega} + \lambda e^{-ih\omega} - \lambda - 1)(t) * u(t)$$

$$(Tu)(t) = \frac{u(t+h)-u(t)}{h} - \lambda\left(\frac{u(t)-u(t-h)}{h}\right)$$

Table of Contents

Motivation

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

Motivation

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Define**:

$$\mathbf{v}_n = (\mathbf{u}_n - \mathbf{u}_{n-1})/h$$

Motivation

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Define**:

$$\mathbf{v}_n = (\mathbf{u}_n - \mathbf{u}_{n-1})/h$$

- **Two-step**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda\mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda\mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned} \tag{2}$$

Motivation

- **Recall** (1):

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \lambda(\mathbf{u}_n - \mathbf{u}_{n-1}) - h\nabla\Phi(\mathbf{u}_n + a(\mathbf{u}_n - \mathbf{u}_{n-1}))$$

- **Define**:

$$\mathbf{v}_n = (\mathbf{u}_n - \mathbf{u}_{n-1})/h$$

- **Two-step**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda\mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda\mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned} \tag{2}$$

- **Notice** $h = 0$:

$$\begin{aligned} \mathbf{u}_n &= \mathbf{u}_0 \\ \mathbf{v}_n &\to -(1-\lambda)^{-1}\nabla\Phi(\mathbf{u}_0) \end{aligned}$$

Motivation

- **Define**:

$$\mathbf{v}_n = (\mathbf{u}_n - \mathbf{u}_{n-1})/h$$

- **Two-step**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda\mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda\mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned} \tag{2}$$

- **Notice** $h = 0$:

$$\begin{aligned} \mathbf{u}_n &= \mathbf{u}_0 \\ \mathbf{v}_n &\to -(1-\lambda)^{-1}\nabla\Phi(\mathbf{u}_0) \end{aligned}$$

- **Invariant Manifold**: small perturbation to

$$\mathbf{v}_n = -(1-\lambda)^{-1}\nabla\Phi(\mathbf{u}_n)$$

Motivation

- **Two-step**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda \mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned} \tag{2}$$

- **Notice** $h = 0$:

$$\begin{aligned} \mathbf{u}_n &= \mathbf{u}_0 \\ \mathbf{v}_n &\to -(1-\lambda)^{-1}\nabla\Phi(\mathbf{u}_0) \end{aligned}$$

- **Invariant Manifold**: small perturbation to

$$\mathbf{v}_n = -(1-\lambda)^{-1}\nabla\Phi(\mathbf{u}_n)$$

seek $g : \mathbb{R}^d \to \mathbb{R}^d$ s.t. the manifold

$$\mathbf{v} = \underbrace{(1-\lambda)^{-1}}_{\bar{\lambda}}\underbrace{(-\nabla\Phi(\mathbf{u}))}_{f(\mathbf{u})} + hg(\mathbf{u})$$

is invariant under the dynamic:

$$\mathbf{v}_n = \bar{\lambda} f(\mathbf{u}_n) + hg(\mathbf{u}_n) \iff \mathbf{v}_{n+1} = \bar{\lambda} f(\mathbf{u}_{n+1}) + hg(\mathbf{u}_{n+1})$$

Existance of Invariant Manifold

Theorem

Fix $\lambda \in (0,1)$ *and let,* $\mathbf{u}_n, \mathbf{v}_n$*, for* $n = 0, 1, 2, \ldots$*, be given by* (2)*. Then there is a* $\tau > 0$ *such that for any* $h \in [0, \tau)$ *there exists a unique* $g \in \Gamma(\gamma, \delta)$ *such that*

$$\mathbf{v}_n = \bar{\lambda} f(\mathbf{u}_n) + hg(\mathbf{u}_n) \Longleftrightarrow \mathbf{v}_{n+1} = \bar{\lambda} f(\mathbf{u}_{n+1}) + hg(\mathbf{u}_{n+1}).$$

Furthermore, the manifold is exponentially attractive,

$$|\mathbf{v}_n - \bar{\lambda} f(\mathbf{u}_n) - hg(\mathbf{u}_n)| \leq (\lambda + h^2 \lambda \delta)^n |\mathbf{v}_0 - \bar{\lambda} f(\mathbf{u}_0) - hg(\mathbf{u}_0)|$$

where $\lambda + h^2 \lambda \delta < 1$*.*

Intuition

- **Recall**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda \mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda \mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned}$$

$$\mathbf{v}_n = \bar{\lambda} f(\mathbf{u}_n) + hg(\mathbf{u}_n) \iff \mathbf{v}_{n+1} = \bar{\lambda} f(\mathbf{u}_{n+1}) + hg(\mathbf{u}_{n+1})$$

Intuition

- **Recall**:

$$\begin{aligned} \mathbf{u}_{n+1} &= \mathbf{u}_n + h\lambda\mathbf{v}_n - h\nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \\ \mathbf{v}_{n+1} &= \lambda\mathbf{v}_n - \nabla\Phi(\mathbf{u}_n + ha\mathbf{v}_n) \end{aligned}$$

$$\mathbf{v}_n = \bar{\lambda} f(\mathbf{u}_n) + hg(\mathbf{u}_n) \iff \mathbf{v}_{n+1} = \bar{\lambda} f(\mathbf{u}_{n+1}) + hg(\mathbf{u}_{n+1})$$

- **Define**: $T : C(\mathbb{R}^d;\mathbb{R}^d) \to C(\mathbb{R}^d;\mathbb{R}^d)$ by

$$\begin{aligned} p &= \xi + h\lambda(\bar{\lambda} f(\xi) + hg(\xi)) + hf(\xi + ha(\bar{\lambda} f(\xi) + hg(\xi))) \\ \bar{\lambda} f(p) + h(Tg)(p) &= \lambda(\bar{\lambda} f(\xi) + hg(\xi)) + f(\xi + ha(\bar{\lambda} f(\xi) + hg(\xi))) \end{aligned}$$

fixed point of $g \mapsto Tg$ gives invariant manifold.

Intuition

- **Define**: $T : C(\mathbb{R}^d; \mathbb{R}^d) \to C(\mathbb{R}^d; \mathbb{R}^d)$ by

$$\begin{aligned} p &= \xi + h\lambda(\bar{\lambda} f(\xi) + hg(\xi)) + hf(\xi + ha(\bar{\lambda} f(\xi) + hg(\xi))) \\ \bar{\lambda} f(p) + h(Tg)(p) &= \lambda(\bar{\lambda} f(\xi) + hg(\xi)) + f(\xi + ha(\bar{\lambda} f(\xi) + hg(\xi))) \end{aligned}$$

fixed point of $g \mapsto Tg$ gives the invariant manifold.

- **Taylor Expansion**:

$$\begin{aligned} p &= \xi + hz_g(\xi) \\ (Tg)(p) &= \lambda g(\xi) + aI_g^{(1)}(\xi) - \bar{\lambda} I_g^{(2)}(\xi) \end{aligned}$$

where

$$\begin{aligned} w_g(\xi) &= \bar{\lambda} f(\xi) + hg(\xi) \\ z_g(\xi) &= \lambda w_g(\xi) + f(\xi + haw_g(\xi)) \end{aligned}$$

Intuition

- **Taylor Expansion**:

$$\begin{aligned} p &= \xi + hz_g(\xi) \\ (Tg)(p) &= \lambda g(\xi) + aI_g^{(1)}(\xi) - \bar{\lambda} I_g^{(2)}(\xi) \end{aligned}$$

where

$$\begin{aligned} w_g(\xi) &= \bar{\lambda} f(\xi) + hg(\xi) \\ z_g(\xi) &= \lambda w_g(\xi) + f(\xi + haw_g(\xi)) \end{aligned}$$

- **Set**: $h = 0$

$$g(p) = \bar{\lambda}^2 (a - \lambda) Df(p) f(p)$$

Intuition

- **Taylor Expansion**:

$$p = \xi + hz_g(\xi)$$
$$(Tg)(p) = \lambda g(\xi) + aI_g^{(1)}(\xi) - \bar{\lambda} I_g^{(2)}(\xi)$$

where

$$w_g(\xi) = \bar{\lambda} f(\xi) + hg(\xi)$$
$$z_g(\xi) = \lambda w_g(\xi) + f(\xi + haw_g(\xi))$$

- **Set**: $h = 0$

$$g(p) = \bar{\lambda}^2 (a - \lambda) Df(p) f(p)$$
$$g(p) = \frac{1}{2} \bar{\lambda}^2 (a - \lambda) \nabla(|\nabla \Phi(p)|^2) \qquad (f(p) = -\nabla \Phi(p))$$

New Modified Equation

- **Dynamic**:
$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi_h(u)$$
where
$$\Phi_h(u) = \Phi(u) + \frac{1}{2}hc|\nabla\Phi(u)|^2$$

Convergence

Theorem

Suppose the assumptions of the existance theorem hold. Let $u \in C^3([0,\infty);\mathbb{R}^d)$ *be the solution to*

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi_h(u)$$

with $c = \bar{\lambda}(\bar{\lambda} - a + 1)$. *For* $n = 0, 1, 2, \ldots$ *let* $\mathbf{u}_n$ *be given by* (1) *and define* $u_n = u(nh)$. *Then for any initial data on the invariant manifold and any* $T \geq 0$, *there is a constant* $C = C(T)$ *such that*

$$\sup_{0 \leq nh \leq T} |u_n - \mathbf{u}_n| \leq Ch^2$$

Conclusion

- Gradient flow:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\Phi(u)$$

- Viscosity equation:

$$h\alpha\frac{d^2u}{dt^2} + (1-\lambda)\frac{du}{dt} + \nabla\Phi(u) = 0$$

- Perturbed gradient flow:

$$\frac{du}{dt} = -(1-\lambda)^{-1}\nabla\left(\Phi(u) + \frac{1}{2}hc|\nabla\Phi(u)|^2\right)$$