Solving Hamilton-Jacobi Bellman equations on Adaptive Sparse Grids

Jochen Garcke
joint work with Ilja Kalmykov, Axel Kröner, Steffen Rutscheidt

UNIVERSITÄT BONN
Institute for Numerical Simulation

Fraunhofer
SCAI
Department Numerical Data-Driven Prediction

Fraunhofer Center for Machine Learning

Fraunhofer-Gesellschaft

The Largest Organization for Applied Research in Europe

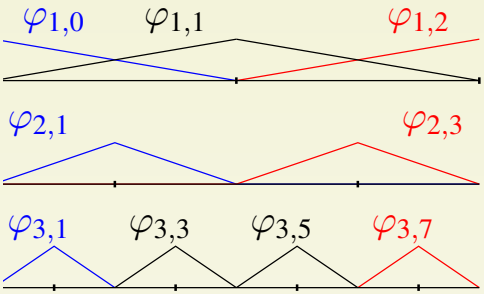
- Application-oriented research for immediate benefit to the economy and to the benefit of society
- **72 institutes** and research units
- **26,600 staff**
- **€2.6 billion annual research budget** totaling. Of this sum, more than €2.2 billion is generated through contract research
  - Roughly two thirds of this sum is generated through contract research on behalf of industry and publicly funded research projects
  - Roughly one third is contributed by the German federal and Countries governments in the form of base funding

© Fraunhofer

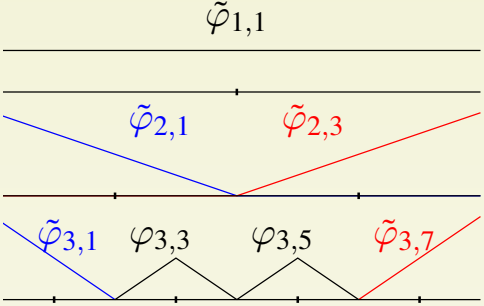
Outline

1. Sparse Grids
2. Semi-Lagrangian Scheme for Optimal control
3. Higher Order Runge-Kutta Methods
4. Finite Differences on Sparse Grids for Economic Problems

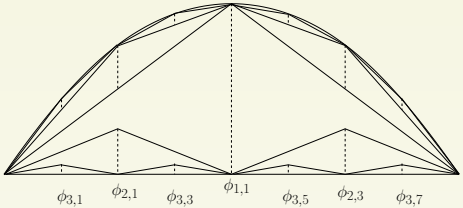
Hierarchical Basis

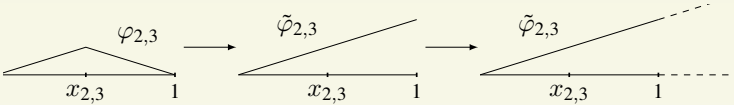


(a) hierarchical basis $V_3 = W_3 \bigoplus W_2 \bigoplus V_1$



(b) fold-out hierarchical basis

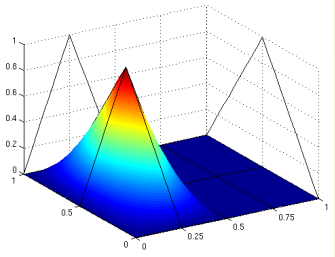




Hier. Basis Functions in Higher Dimensions

- d-dimensional piecewise d-linear functions

$$\phi_{\underline{l},\underline{j}}(\underline{x}) \;:=\; \prod_{t=1}^{d} \phi_{l_t,j_t}(x_t)$$



- hierarchical difference space $W_{\underline{l}}$ ($\underline{e}_t$ is t-th unit vector)

$$W_{\underline{l}} \;:=\; V_{\underline{l}} \;\setminus\; \bigcup_{t=1}^{d} V_{\underline{l}-\underline{e}_t},$$

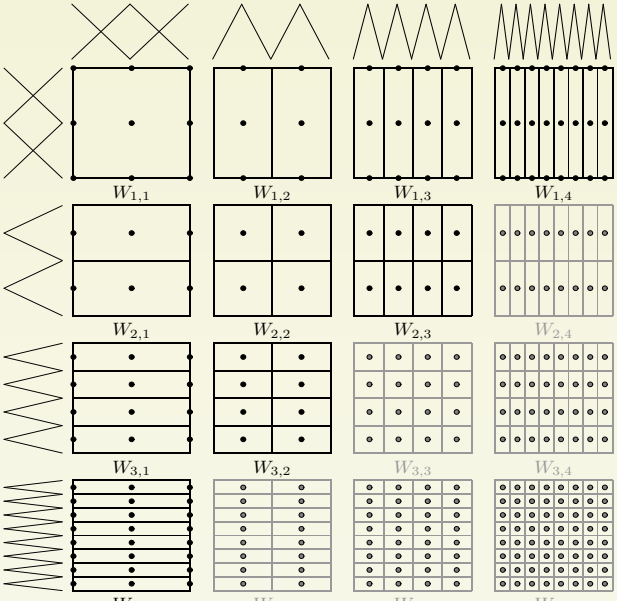
- hier. diff. space represented by $W_{\underline{l}} = \operatorname{span}\{\phi_{\underline{l},\underline{j}} \,|\, \underline{j} \in \mathrm{B}_{\underline{l}}\}$ with

$$\mathrm{B}_{\underline{l}} \;:=\; \left\{ \underline{j} \in \mathbb{N}^d \;\middle|\; \begin{array}{lll} j_t = 1, \ldots, 2^{l_t} - 1, & j_t \text{ odd}, & t = 1, \ldots, d, \text{ if } l_t > 1, \\ j_t = 0, 1, 2, & & t = 1, \ldots, d, \text{ if } l_t = 1 \end{array} \right\}.$$

- full grid space in hierarchical basis

$$V_n^s \;:=\; \bigoplus_{|\underline{l}|_\infty \le n} W_{\underline{l}}$$

Hierarchical Subspaces $W_{\underline{l}}$



Sparse Grids

- we define the sparse grid function space $V_n^s \subset V_n$ as

$$V_n^s := \bigoplus_{|\underline{l}|_1 \leq n+d-1} W_{\underline{l}}$$

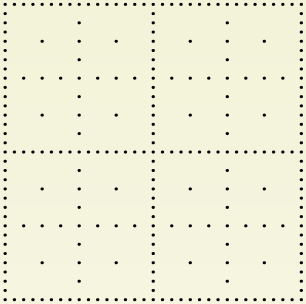
- every $f \in V_n^s$ can now be represented as

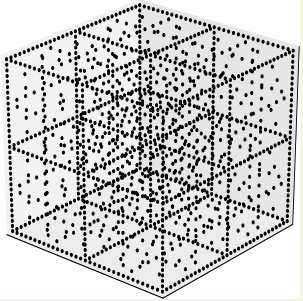
$$f_n^s(\underline{x}) = \sum_{|\underline{l}|_1 \leq n+d-1} \sum_{\underline{j} \in \mathrm{B}_{\underline{l}}} \alpha_{\underline{l},\underline{j}} \phi_{\underline{l},\underline{j}}(\underline{x})$$

- approximation property in H^2_{mix}

$$||f - f_n^s||_2 = O(h_n^2 \log(h_n^{-1})^{d-1})$$

- sparse grid needs $O(h_n^{-1}(\log(h_n^{-1}))^{d-1})$ points





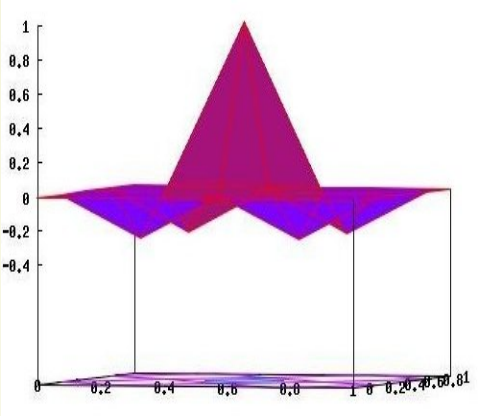
Sparse Tensor Decomposition

- sparse grids introduced by Zenger, 1991
- original idea can be identified in work of Smolyak, 1963
- other related approaches using construction exists as well
- overview from Bungartz and Griebel in Acta Numerica 2004
- tutorial introduction in G. 2013
- truncated tensor construction can be used for "any" multi-resolution representation
  - polynomials
  - B-splines
  - wavelets
  - Fourier series (hyperbolic cross approximation)
  - multi-level Monte Carlo (different resolutions for random variables as well as discretizations)
  - time vs. space discretization
- sparsity pattern can be generalized from simplex or can be dimension-adaptive
- SG++-library from Dirk Pflüger (U. Stuttgart) with Matlab and Python-bindings
- in UQ-context and for numerical integration further libraries exist

Problems with Sparse Grids for HJB: Monotonicity

- interpolation of peaked Gaussian fct. with sparse grid of level $n = 2$

$$f(x_1, x_2) := \exp\left(-100(x_1 - 0.5)^2\right) * \exp\left(-100(x_2 - 0.5)^2\right)$$



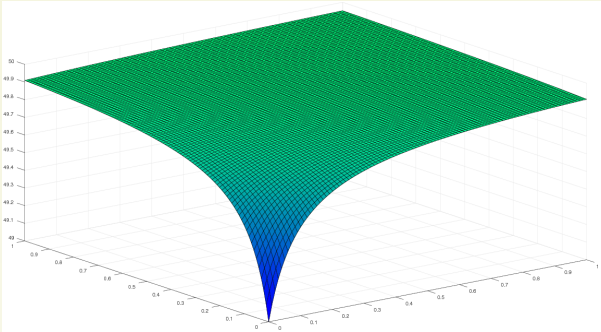
- sparse grid interpolation does not preserve positivity

Problems with Sparse Grids for HJB: Monotonicity

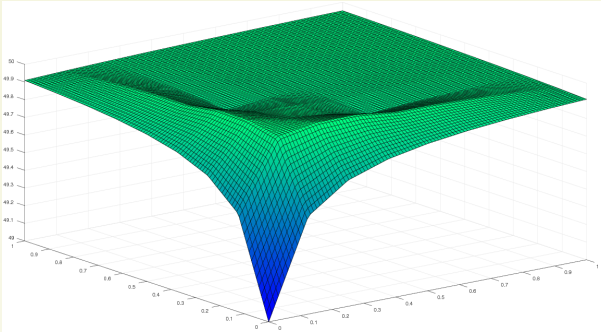
- problem persists for monotonically increasing (strictly) concave functions, e.g.

$$f(x,y) = \frac{-1}{1+10x+10y} + 50.$$

- function is similar to value functions arising for heterogenous agent models in economics



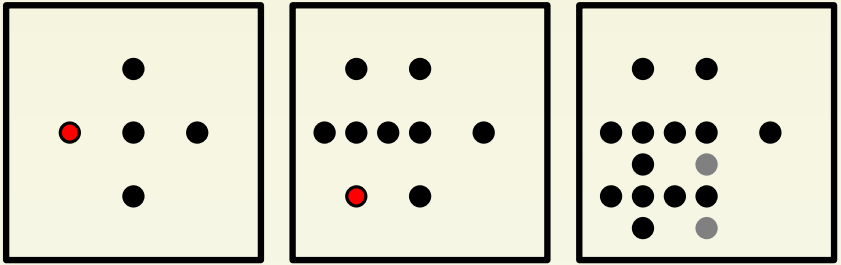
function plot of f



interpolation plot of f

Spatially Adaptive Sparse Grids

- spatially adaptive grids can be used when functions
  - do not fulfil smoothness condition or
  - strongly vary due locally large derivatives
- adaptivity helps to some degree to cope with non-monotonicity
  - refine one grid point by creating all children (middle)
  - to keep grid consistent, missing parents are created (right)



- usually hierarchical surplus $\alpha_{\underline{l},\underline{j}}$ is used as refinement indicator
- in case of monotonically increasing concave functions one could check for negative derivatives

Optimal Feedback Control

- optimal feedback control of a dynamical system

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} J(u) = \int_0^T l(y(t), u(t))\,\mathrm{d}t, \quad \text{s.t.} \\ \dot{y}(t) = f(y(t), u(t)), \quad t > 0 \\ y(0) = y_0 \end{cases}$$

  - STATE $y(t) \in \mathbb{R}^d$, initial state $y_0 \in \mathbb{R}^d$
  - CONTROL $u(t) \in U^m \subset \mathbb{R}^m$ (often called action)
  - Lipschitz continuous DYNAMICS $f\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$
  - running COST with polynomial growth $l\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$
  - set of ADMISSIBLE CONTROLS
    $$\mathcal{U}_{ad} = \{u \in L^2([0,T]; U^m) \mid U^m \subset \mathbb{R}^m \text{ compact}\}$$
- aim: FEEDBACK LAW $\qquad u^* = \mathbf{K}(t, y^*(t))$

Basic Sparse Grid Semi-Lagrangian Scheme

evaluate for $\underline{x} \in Q_{\mathcal{I}}$, $\Delta t > 0$, $K = T/\Delta t$, $k = K-1, \dots, 0$,

$$\begin{cases} v^k(\underline{x}) = \min_{u \in U} \left(\Delta t l(\underline{x}, u) + v^{k+1}(y_{\underline{x}}(\Delta t)) \right), \\ v^k(\underline{x}) = 0 \end{cases}$$

$y_{\underline{x}}(\Delta t)$ state obtained by time discretization scheme from $\underline{x}$

minimization either by comparing v^{k+1} values over finite set, nonlinear optimization, or gradient ∇v

Algorithm 1: Adaptive SL-SG scheme

Data: refinement constant ε, coarsening constant η

Result: sequence of adaptive sparse grid solutions $v_k \in V_{\mathcal{I}(k)}$

initialize $\mathcal{I}(K)$

for $k = K-1, \dots, 0$ **do** ▷ ITERATE IN TIME WITH $\Delta t = T/K$

  initialize $\mathcal{I}(k-1)$ with $\mathcal{I}(k)$

  adaptively interpolate $\min_{u \in U} \left(v_k(y_{\underline{x}}(\Delta t)) + \Delta t l(\underline{x}, u) \right)$ ▷ COMPUTE v_{k-1}

  coarsen $v_{k-1} \in V_{\mathcal{I}(k-1)}$

see Bokanowski, G., Griebel, and Klompmaker (2013)

Example 2D: Simplified Semi-Discrete Wave Equation

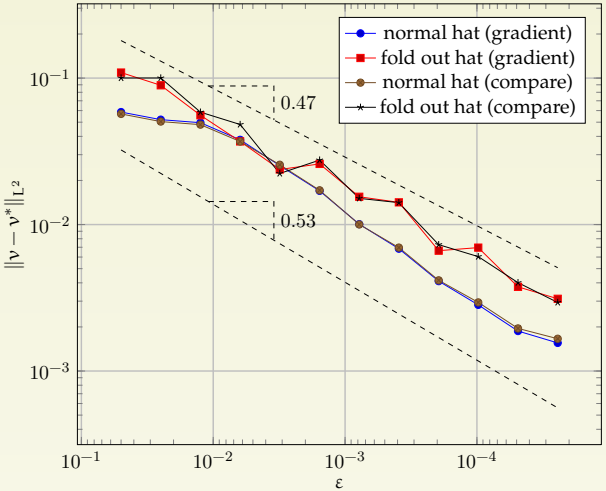
- semi-discrete PDE control problem (following Kröner, Kunisch, Zidani (2015))
- formulate as first order system in time with $y^1 = y$, $y^2 = \dot{y}$
- dynamics $f(x, u) = \mathcal{A}x + \mathcal{B}u$
- with running cost $l(x, u) := \beta_x x_1^T M x_1 + \beta_u u^T u$
- Hamilton-Jacobi Bellman equation with dimension $2d$
- first consider a simplified example based on harmonic oscillator

$$\beta_x = 2,\ \beta_u = 0.1, \quad T = 1,\ \Delta t = 0.01,$$
$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

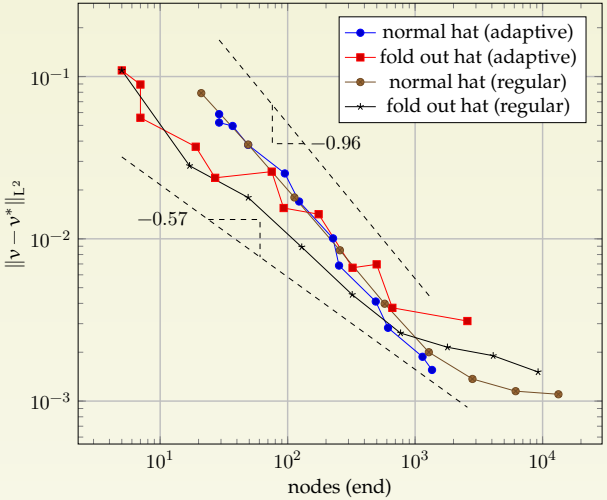
initial data $x \in \mathbb{R}^2$, domain $Q = [-1, 1]^2$, $U = [-3.5, 3.5]$.

- reference solution computed with a higher order FD code on very fine mesh
- further details in G., Kröner (2017)

Example 2D: Convergence of value function

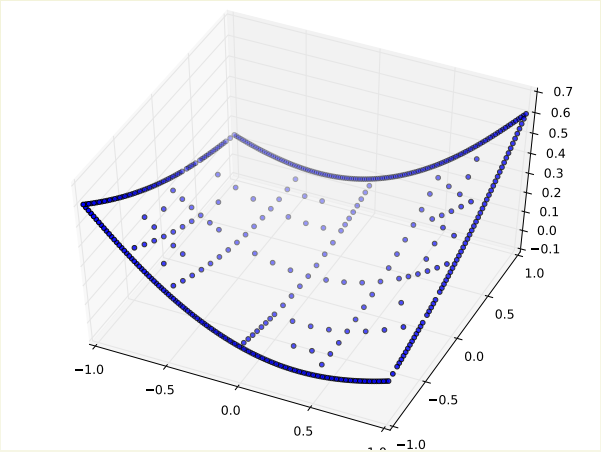


(e) L_2 error in the value function vs. ε

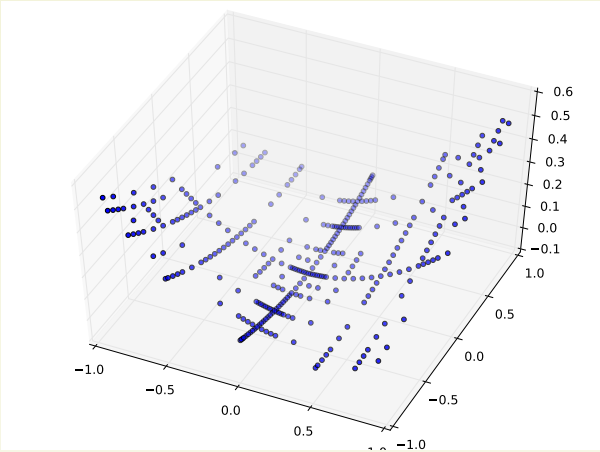


(f) L_2 error in the value function vs. nodes

Example 2D: Value function



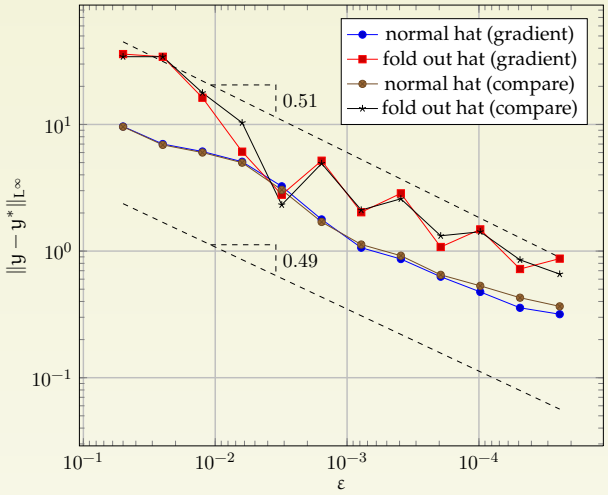
(g) $\varepsilon = 1.95_{-4}$



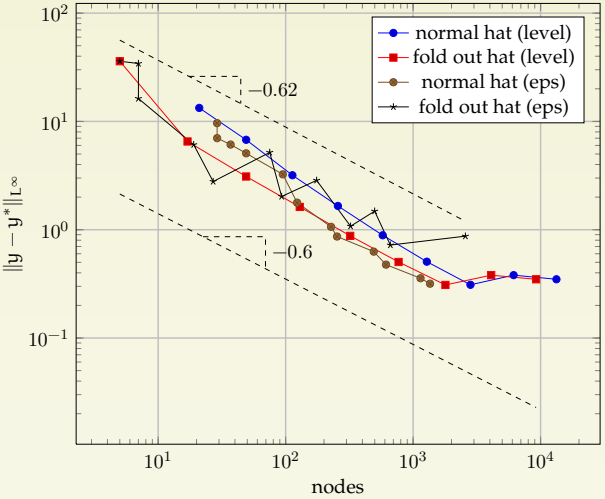
(h) $\varepsilon = 1.95_{-4}$

Figure: adaptive sparse grid with normal and fold out hat functions

Example 2D: Convergence in the Trajectory

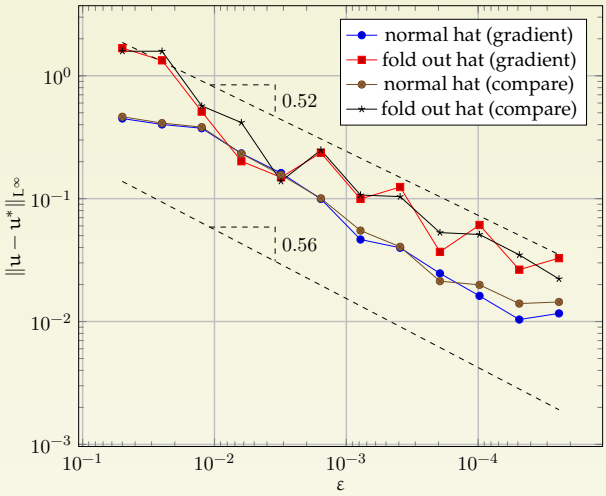


(a) L_∞ error in the trajectory vs. ε

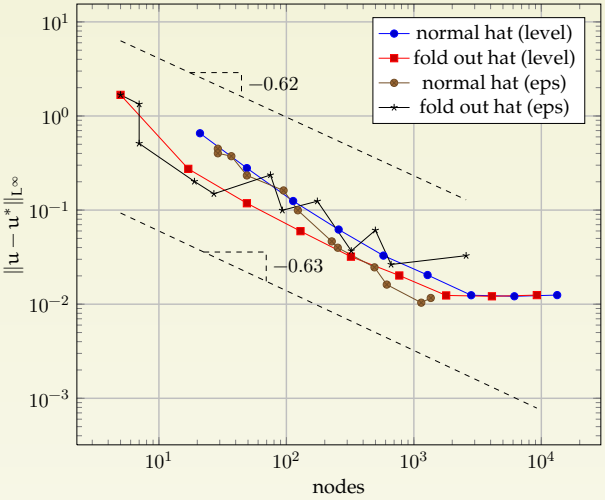


(b) L_∞ error in the trajectory vs. nodes

Example 2D: Convergence in the Control



(c) L_∞ error in the control vs. ε



(d) L_∞ error in the control vs. nodes

Example: Semi-Discrete Wave Equation (4D)

- now wave equation as first order system in time

$$\dot{y}(t) = f_{\mathrm{w}}(y(t), u(t)), \quad t > 0, \quad y(0) = y_0$$

- with dynamics

$$f_{\mathrm{w}} \colon \mathbb{R}^{2d} \times \mathbb{R}^m \to \mathbb{R}, \quad f_{\mathrm{w}}(x, u) := \mathcal{A}x + \mathcal{B}u$$

where

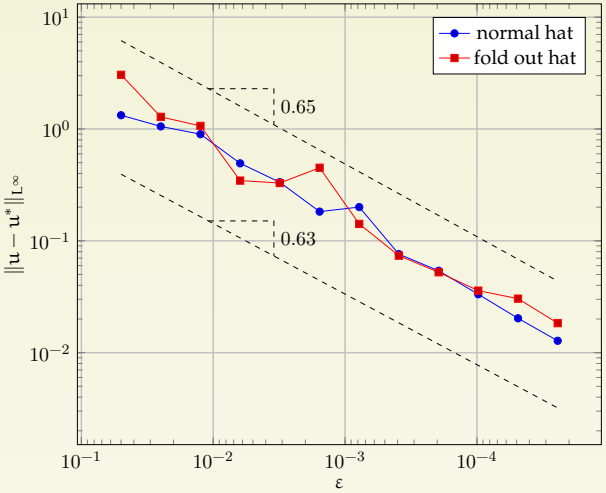
$$\mathcal{A} := \begin{pmatrix} 0 & I_d \\ -cM^{-1}A & 0 \end{pmatrix}, \quad \mathcal{B} := \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad b \in \mathbb{R}^{m \times d}, \quad y_0 \in \mathbb{R}^{2d}$$

- cost $\quad l_w(x, u) := \beta_x x_1^T M x_1 + \beta_u u^T u$
- consider setup

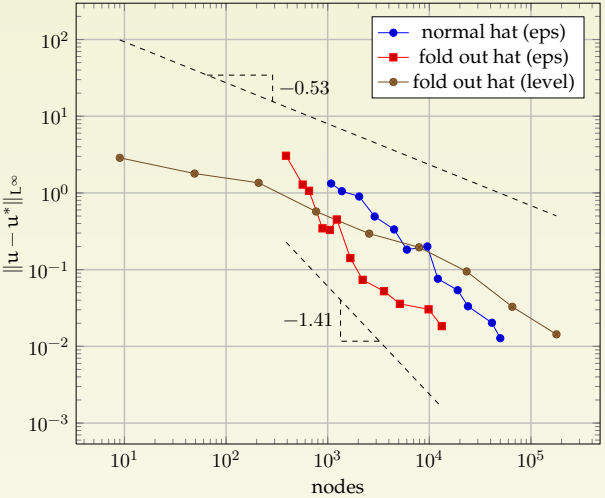
$$\beta_x = 2, \quad \beta_u = 0.1, \quad T = 4, \quad \Delta t = 0.01, \quad c = 0.05.$$

- compute reference trajectories in state space and control space using a Riccati approach

Example: Convergence in the Control (4D)

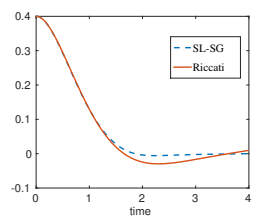


(e) L_∞ error in the control vs. ε

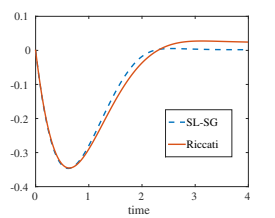


(f) L_∞ error in the control vs. nodes

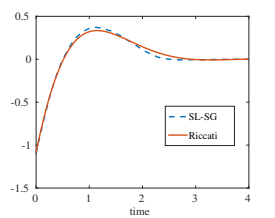
Example: semi-discrete wave equation (4D)



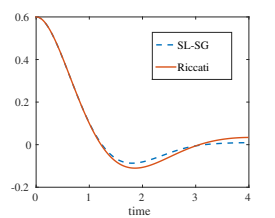
(g) y_1



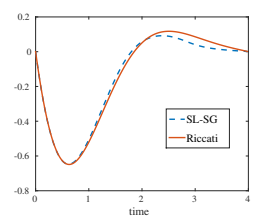
(h) y_3



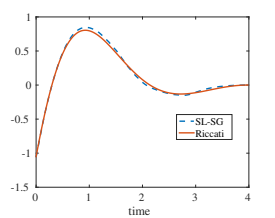
(i) u_1



(j) y_2

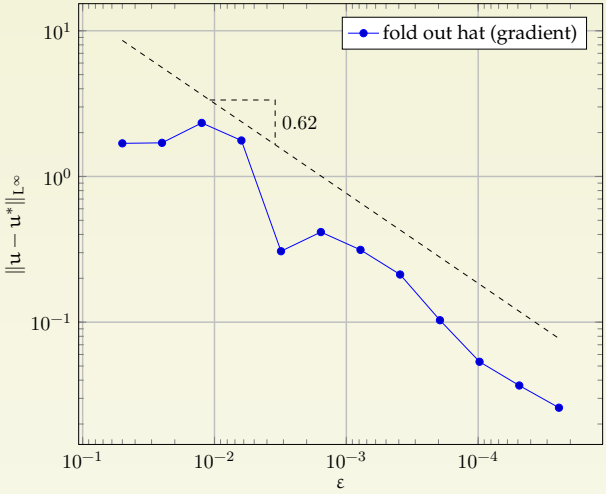


(k) y_4

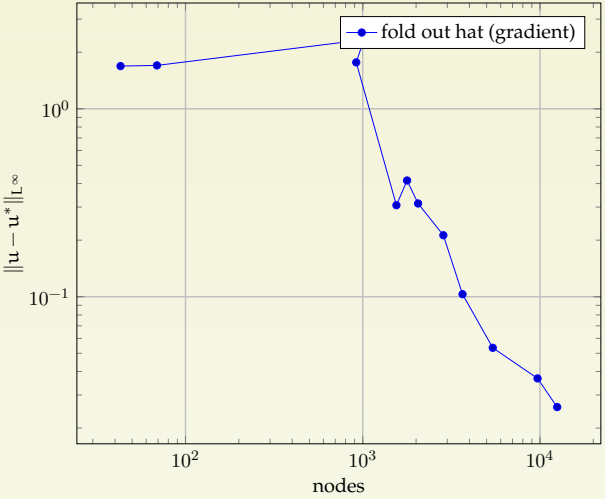


(l) u_2

Example: semi-discrete wave equation (6D)



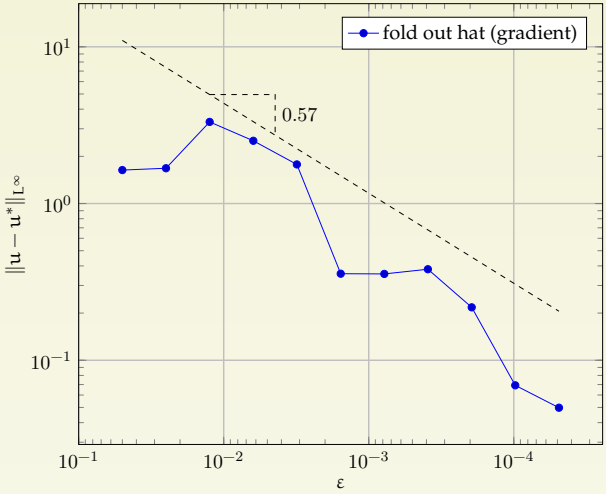
(a) error in the control vs. ε



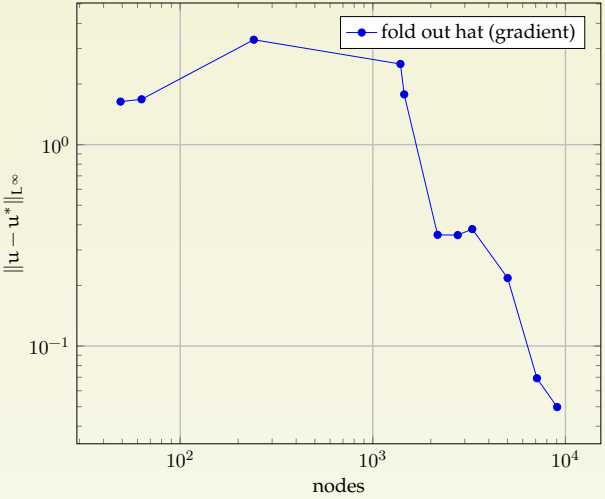
(b) error in the control vs. nodes

need to decrease time step to $\Delta t = 0.0025$ (larger entries in stiffness matrix)

Example: semi-discrete wave equation (8D)



(c) error in the control vs. ε



(d) error in the control vs. nodes

need to decrease time step to $\Delta t = 0.00125$ (larger entries in stiffness matrix)

Higher Order Runge-Kutta Methods

- look at control formulation

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} J(u) = \int_0^T l(y(t), u(t))\,\mathrm{d}t, \quad \text{s.t.} \\ \qquad \dot{y}(t) = f(y(t), u(t)), \quad t > 0, y(0) \quad = y_0 \end{cases}$$

- we need time discretisation for $y(t)$ and
- quadrature rule to evaluate $\min_{u \in \mathcal{U}_{ad}} J(u)$
- with higher order RK-scheme this could look like (misusing notation)

$$v^k(\underline{x}) = \min_{u_1, u_2, u_3, \ldots} \left(\Delta t \sum_i c_i l(\underline{x}^{\tau_i}, u_i) + v^{k+1}(\hat{y}_{\underline{x}}(\Delta t)) \right)$$

  where $\hat{y}_{\underline{x}}$ takes into account actions u_i and RK-scheme with intermediate steps τ_i
- therefore RK4 would have $\mathcal{O}(d_c^4)$ complexity if complexity for computing one control is d_c
- see also Falcone, Ferretti (1994)

Structure Preserving Runge-Kutta Methods

- observe that for explicit or diagonally implicit RK schemes, the last element u_s^n of discrete control vector affects only the last RK-stage k_s
- formulate condition on RK-coeff., s.t. control optimization is applied separately to each stage
- idea is to mimic Dynamic Programming property within single step of semi-Lagrangian scheme
- class of RK methods fulfilling this: diagonally implicit symplectic Runge-Kutta (DISRK) schemes
- originally developed for long time integration of Hamiltonian systems
- "re-use" implicit collocation RK-scheme for quadrature
- DISRK are equivalent to composition of implicit midpoint schemes $\Psi_h = \Phi_{\gamma_s \tau} \circ \cdots \circ \Phi_{\gamma_2 \tau} \circ \Phi_{\gamma_1 \tau}$
- we can construct a SL scheme, which has $\mathcal{O}(d_c \cdot s)$ complexity for the minimization problem
- for example DISRK_5 is a method of order 4 using 5 steps

$$\gamma_1 = \gamma_2 = \gamma_4 = \gamma_5 = \frac{1}{4 - 4^{1/(p+1)}}, \quad \gamma_3 = -\frac{4^{1/(p+1)}}{4 - 4^{1/(p+1)}} \tag{1}$$

- observe for $s > 1$ DISRK goes backwards in time for some steps !
- details in G., Kalmykov (2018)

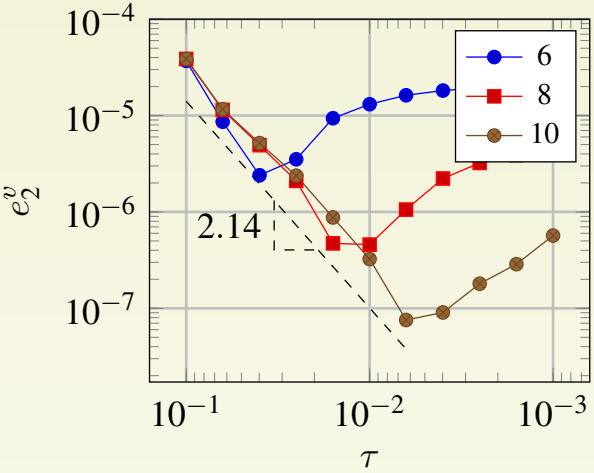
Harmonic Oszillator as Simplified Semi-Discrete Wave - Revisited

- consider again the simplified example based on harmonic oscillator

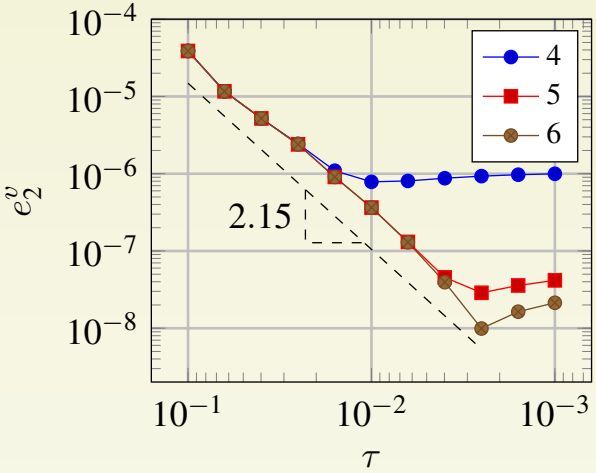
$$\beta_x = 2,\ \beta_u = 0.1, \quad T = 0.1,\ \Delta t = 0.01,$$
$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

  initial data $x \in \mathbb{R}^2$, domain $Q = [0,1]^2$, $U = [-3.5, 3.5]$.
- control space discretization is denoted by A_Δ
- use sparse grids with linear and 3rd order B-Splines, both with fold out ansatz (SG++-library)
- equation for stage k of Implicit Mipoint Rule is solved with standard Newton algorithm
- obtain a reference solution from continuous time Riccati differential equation
- computation of error estimates is performed on a domain $\Omega_{ref} = [0.25, 0.75]^2$

Simplified Semi-Discrete Wave - Revisited with DISRK1

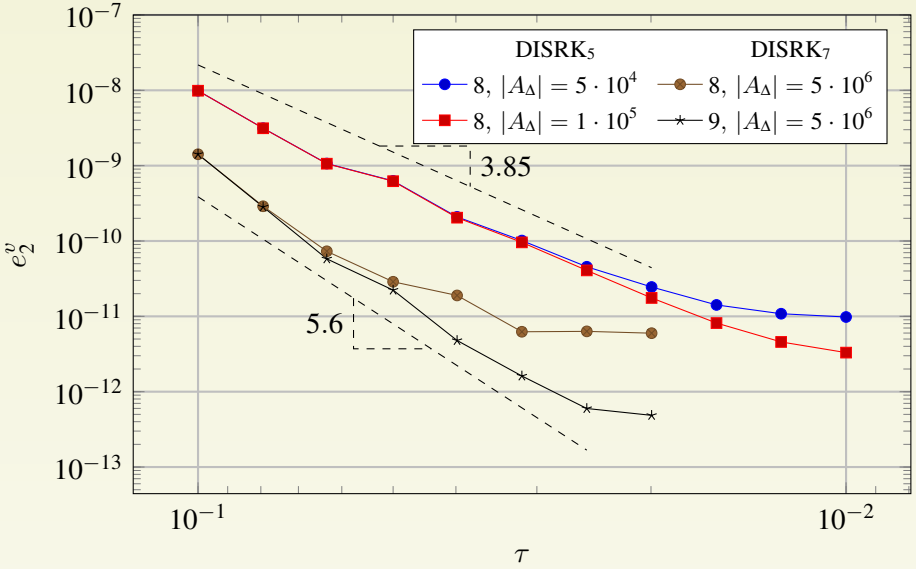


(e) implicit midpoint, lin. sparse grids



(f) implicit midpoint, sparse grids 3rd order B-splines

Simplified Semi-Discrete Wave



Finite Differences on Sparse Grids

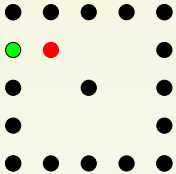
- finite differences on sparse grids were introduced by [Griebel.Schiekofer:1998]
- there, finite difference operators are a composition of three partial operators
  1. a basis transformation from nodal to hierarchical basis in all dimensions but the dimension j in which we aim to use the finite difference stencil
  2. application of a finite difference stencil in dimension j, where mesh size is given as local step size to the neighboring grid point in dimension j
  3. a basis transformation from hierarchical to nodal basis in all dimensions but dimension j
- consistency proofs can be given for elliptic PDE model problems
- we investigate an alternative from [Ahn.2017], where additional ghost nodes are used
- instead of using function values on the sparse grid points, one interpolates on ghost nodes
- therefore, one does not need to use specific basis transformations and one can simply take any sparse grid library, such as SG++ [Pflüger:2010]

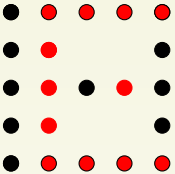
Finite Differences on Sparse Grids Using Ghost Nodes

- we define the *ghost node step size* h_{g_j} in dimension j, $1 \leq j \leq n$, for a grid point $x_{\mathbf{l},\mathbf{i}}$ by $h_{g_j} := 2^{-k_j}$ where k_j denotes the maximal level used in dimension j
- for a grid point $x_{\mathbf{l},\mathbf{i}}$ in which we aim to compute the finite differences in dimension j, $1 \leq j \leq d$, we define the corresponding *forward difference ghost node* by

$$g^{F,j}_{\mathbf{l},\mathbf{i}} := (x_{l_1,i_1}, \ldots, x_{l_j,i_j} + h_{g_j}, \ldots, x_{l_d,i_d})$$

- for backward difference corresponding *backward difference ghost nodes* are defined accordingly
- left: ghost node (red) that is used for sparse grid forward FD in x-dimension in green grid point
- right: all forward difference ghost nodes that are used for the sparse grid are drawn in red





Finite Differences on Sparse Grids Using Ghost Nodes

- with interpolation operator on the sparse grid

$$\mathbf{I}_s : \bigoplus_{\mathbf{k}=\mathbf{1}}^{\mathbf{l}} W_{\mathbf{k}} \to V_{\mathbf{l}}.$$

- and interpolation operator for the by h_{g_j} shifted sparse grid (grid of ghost nodes)

$$\mathbf{I}^F_{h_{g_j}} : \bigoplus_{\mathbf{k}=\mathbf{1}}^{\mathbf{l}} W_{\mathbf{k}} \to V_{\mathbf{l}} \quad \text{and} \quad \mathbf{I}^B_{h_{g_j}} : \bigoplus_{\mathbf{k}=\mathbf{1}}^{\mathbf{l}} W_{\mathbf{k}} \to V_{\mathbf{l}}.$$

  where $\mathbf{I}^F_{h_{g_j}}$ is the forward difference and $\mathbf{I}^B_{h_{g_j}}$ the backward difference in dimension j, $1 \leq j \leq d$

- we can defne the *sparse grid backward difference operator* $\tilde{\mathbf{D}}^{S,B}_j$ by

$$\tilde{\mathbf{D}}^{S,B}_j := \mathbf{I}_s - \mathbf{I}^B_{h_{g_j}} : \bigoplus_{\mathbf{k}=\mathbf{1}}^{\mathbf{l}} W_{\mathbf{k}} \to V_{\mathbf{l}}.$$

- for sparse grids in 2D up to level 3 we confirmed the equivalence to [Griebel.Schiekofer:1998]
- an extended version of [Ahn:2017] will cover the relation of the two FD approaches in more detail

Economic Model Problem from [Kaplan.Moll.Violante:2019]

- want to solve the maximization problem

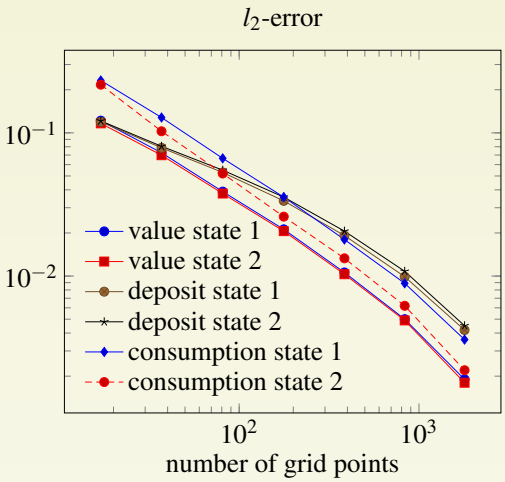
$$\max_{\{c_t,d_t\}_{t\geq 0}} \mathbb{E}_0 \int_0^\infty e^{-\rho t} u(c_t)dt \tag{2}$$

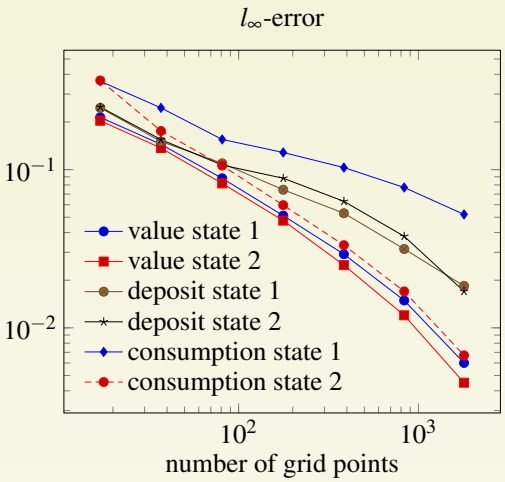
subject to

$$\begin{aligned} \dot{b}_t &= wz_t r^b(b_t)b_t - d_t - \chi(d_t, a_t) - c_t \qquad & \dot{a}_t &= r^a a_t + d_t \\ z_t &= \text{Poisson with intensities } \lambda(z, z') & b_t &\geq b,\ a_t \geq 0. \end{aligned} \tag{3}$$

- b_t denotes liquid assets and a_t illiquid assets, respective returns on these assets are r^b and r^a
- we have consumption c_t, deposits d_t, transaction cost function χ and wage w
- idiosyncratic productivity z_t follows a Poisson process with intensities $\lambda(z, z')$
- Constant Relative Risk Aversion utility function u is strictly convex, strictly monotone increasing
- use POLICY ITERATION, i.e. solve a corresponding linear equation system for fixed policy
- high(er) dimensionality by number of assets and number of skills resulting in productivity
- for details see G., Ruttscheidt (2019)

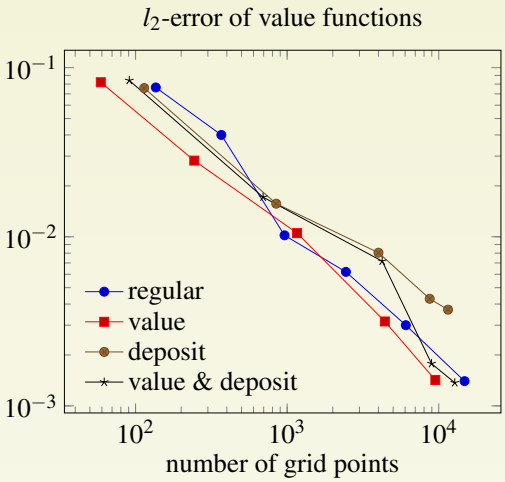
2D model problem

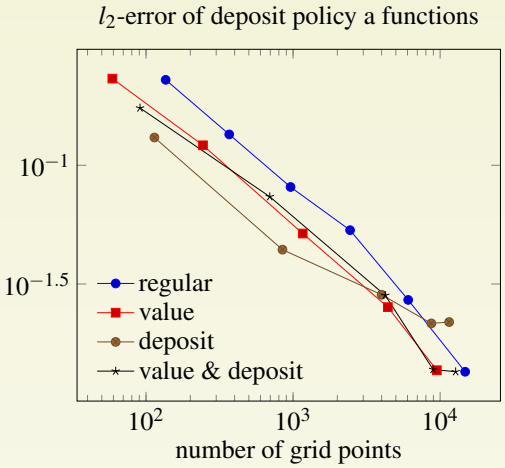




- errors for value function and all policy functions for regular sparse grids of different levels
- reference solution is computed on a 600×600 full grid

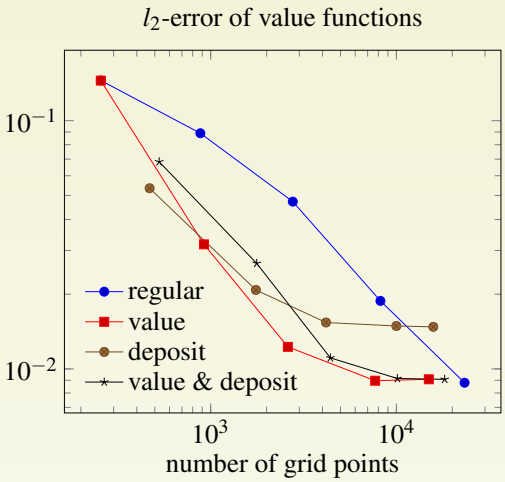
4D model problem

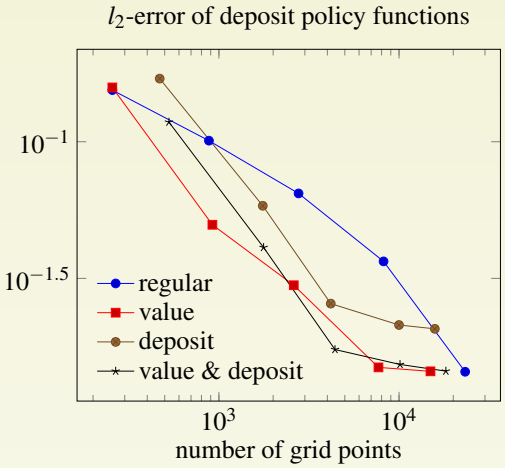




- errors for value function and one deposit policy functions for adaptive sparse grids
- reference solution is on sparse grid of level 8 (largest computable in reasonable time)

6D model problem





- errors for value function and one deposit policy functions for adaptive sparse grids
- reference solution is on sparse grid of level 6 (largest computable in reasonable time)

Conclusion

- use SL-scheme on sparse grids for HJB equations
- investigate model reduction for optimal feedback control PDE problems leading to HJB equations
- suitable for examples in higher dimensions seen up to 8d
- somewhat more complicated function we treated with sparse grids for front propagation models in Bokanowksi, G., Griebel, Klompmaker (2013)
- large scale parallel studies for economic problem in up to 20D by Brumm, Scheidegger (2017)
- higher order time schemes and higher order discretisation can be effective for smooth problems
- sparse grids with B-splines converge "nicer"
  but have higher computing costs due to need to solve a linear equation system
- investigate FD on sparse grids for economic problems using policy iteration

Key References

- O. Bokanowski, J. G., M. Griebel, and I. Klompmaker **An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations**, J. of Sci. Comp., (2013)
- M. Falcone, R. Ferretti, **Discrete time high-order schemes for viscosity solutions of Hamilton-Jacobi-Bellman equations**, Numerische Mathematik. (1994).
- J.G, I. Kalmykov, **Efficient higher order time discretization schemes for Hamilton-Jacobi-Bellman equations based on diagonally implicit symplectic Runge-Kutta methods.** In D. Kalise, K. Kunisch, and Z. Rao, editors, Hamilton-Jacobi-Bellman Equations: Numerical Methods and Applications in Optimal Control, De Gruyter, 2018.
- J. G., A. Kröner, **Suboptimal feedback control of PDEs by solving HJB equations on adaptive sparse grids**, J. of Sci. Comp., (2017).
- J. G. and S. Ruttscheidt, **Finite differences on sparse grids for continuous time heterogeneous agent models.** Proceedings of SGA 2018, (2019)
- A. Kröner, K. Kunisch, H. Zidani, **Optimal feedback control for the undamped wave equation equation**, ESAIM: Control Optim. Calc. Var. (2015).