Bang-bang control as a design principle for heuristic optimization

Aniruddha Bapat[*], Stephen Jordan

arxiv:1812.02746

Outline

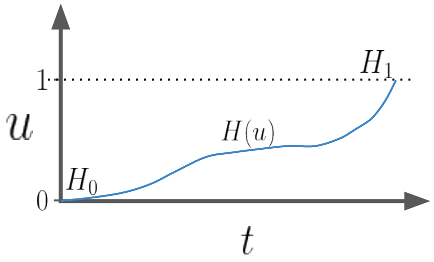
- The optimal control framework (bang-bang, quasistatic)
- Candidate algorithms and instances
- Are bang-bang and quasistatic control polynomially equivalent (No.)
- QAOA1 on symmetric instances
- Caveats to optimal control theory

Heuristic optimization as a control problem

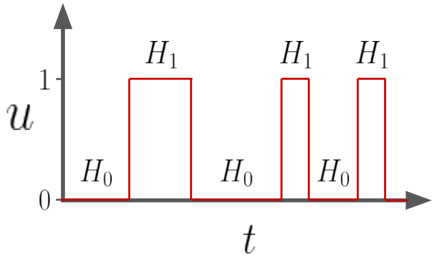
Physically motivated heuristic optimization algorithms seek to prepare a target probability distribution (or state) via a series of controlled moves that guide the evolution.

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -iH(u)|\psi\rangle \quad , \text{ where } \quad H(u) = (1-u)H_0 + uH_1$$

Quasistatic



Bang-bang



The optimal control framework

$$|\psi\rangle = \begin{bmatrix} q_0 \\ q_1 \\ \\ \cdot \\ \cdot \\ \cdot \\ \\ q_N \end{bmatrix}$$

Perform a controlled evolution: $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q} = f(\mathbf{q}, u) \equiv \big[(1-u)H_0 + uH_1\big] \cdot \mathbf{q}, \quad u : [0,T] \to [0,1]$

Goal: Minimize a cost function of the state $\min_u E(\mathbf{q}(T))$ (i.e. find the *optimal* control u^*)

What is a good control strategy?

1. **Quasistatic** - Supported by adiabatic theorems
2. **Bang-bang** - Supported by the Pontryagin Minimum Principle (PMP)

PMP gives necessary conditions for a control to be optimal, in the form of an extremality condition on a classical functional known as the *control Hamiltonian.*

Conjugate momenta: $p_i = \partial_{q_i} E\big|_T$ The control Hamiltonian: $\mathcal{H}(\mathbf{q}, \mathbf{p}, u) := \mathbf{p} \cdot f(\mathbf{q}, u)\ (-\ \ L)$

PMP: u^* satisfies: $\forall t \in [0,T],\ \ \mathcal{H}(\mathbf{q}, \mathbf{p}, u^*) = \min_u \mathcal{H}(\mathbf{q}, \mathbf{p}, u)$

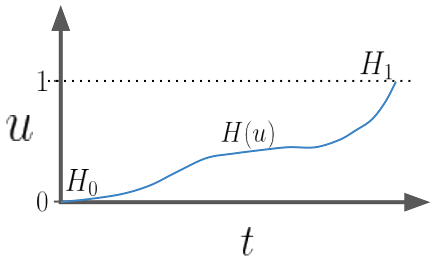
Linear control ⇒ optimal control is necessarily bang-bang i.e. $\mathrm{Range}[u^*] = \{0, 1\}$ *

* *Terms and conditions apply. Please consult your local control theorist.*

Control design

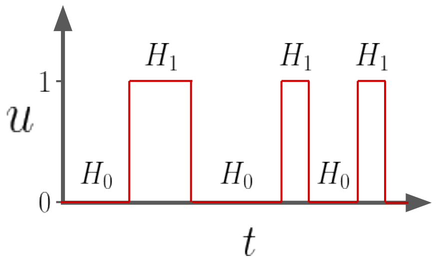
$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -iH(u)|\psi\rangle \quad , \text{ where } \quad H(u) = (1-u)H_0 + uH_1$$

Quasistatic: QAO, QA, SA, SQA, …



?

Bang-bang: QAOA, VQE, …



Q: Are bang-bang control and quasistatic control polynomially equivalent?

The heuristic optimization algorithm alignment chart

QUASISTATIC ⟷ BANG-BANG

QUANTUM ↕ CLASSICAL

Quantum Adiabatic Optimization (QAO)

$$H_0 = -\sum_{i=1}^{n} X_i \qquad H_1 = \sum_{z\in\{0,1\}^n} c(z)|z\rangle\langle z|$$

$$|\psi(u=0)\rangle \xrightarrow{\text{quasistatic}} |\psi(u=1)\rangle$$

Quantum Approximate Optimization Algorithm (QAOA)

$$H_0 = -\sum_{i=1}^{n} X_i \qquad H_1 = \sum_{z\in\{0,1\}^n} c(z)|z\rangle\langle z|$$

$$|+\rangle^{\otimes n} \equiv |\psi_0\rangle \xrightarrow{e^{-i\gamma_1 H_1}} |\psi_1\rangle \xrightarrow{e^{-i\beta_1 H_0}} \cdots \xrightarrow{e^{-i\beta_p H_0}} |\psi_{2p}\rangle$$

$$E(\vec{\beta},\vec{\gamma}) = \langle\vec{\beta},\vec{\gamma}|H_1|\vec{\beta},\vec{\gamma}\rangle$$

Simulated Annealing (SA)

Metropolis-Hastings Monte Carlo with temperature shchedule:

$$\infty \longrightarrow \cdots \longrightarrow T \longrightarrow \cdots \longrightarrow 0$$

and flipping probability:

$$1 \longrightarrow \cdots \min\left\{1, e^{-\Delta_{\text{flip}}V/T}\right\} \cdots \longrightarrow \Theta(\Delta_{\text{flip}}V)$$

Randomized descent ... Diffusion

Bang-bang Simulated Annealing (BBSA)

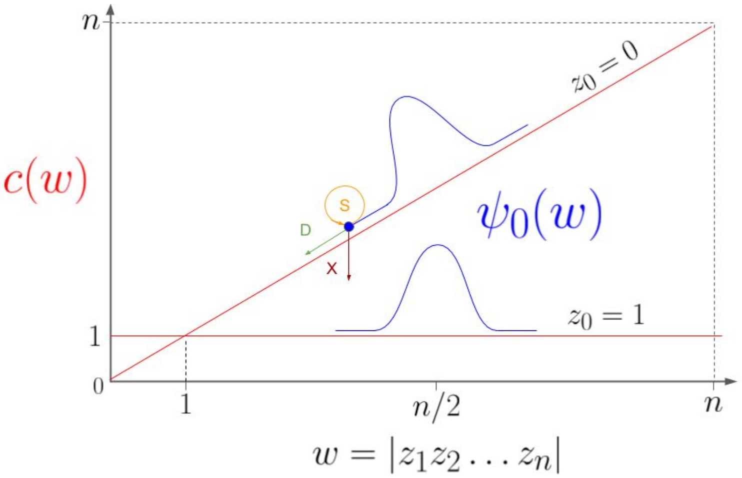
We run MH Monte-Carlo with a bang-bang schedule, i.e., only allowing T=0,∞

This corresponds to alternating periods of randomized descent and diffusion.

The instances

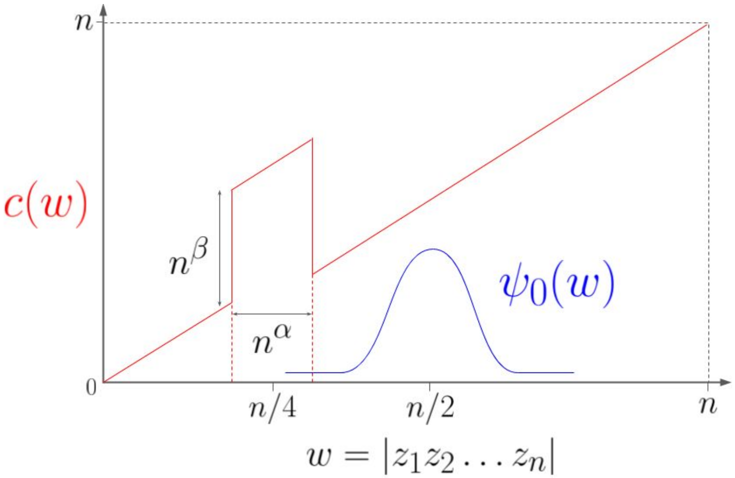
Hamming symmetry: $c(z) \equiv c(w)$, where $w = |z|$ = # of ones in the bit string z

1. Bush of Implications (Bush)



$$c(z_0 z_1 \ldots z_n) = z_0 + \sum_{i=1}^{n} z_i(1 - z_0)$$
$$c(z_0, w) = z_0 + w(1 - z_0)$$

2. Ramp with Spike (Spike)



$$r(w) = w, \quad s(w) = \begin{cases} n^\beta, & \text{if } w \in [\frac{n}{4} - \frac{n^\alpha}{2}, \frac{n}{4} + \frac{n^\alpha}{2}] \\ 0, & \text{otherwise.} \end{cases}$$
$$c(w) = r(w) + s(w)$$

Results

Instance	Annealing-based		Bang-bang	
	QAO	SA	QAOA	BBSA
Bush, $\lambda \geq 1$	$\mathrm{poly}(n)$ [13]	$\exp(n)$ [13]	$O(1)$§ 8.3.2	$\tilde{O}\left(n^{3.5\ldots}\right)$§ 8.2.1
Bush, $\lambda < 1$	$\exp(n)$ [13]	$\exp(n)$ [13]	$O(1)$§ 8.3.2	$\tilde{O}\left(n^{3.5\ldots}\right)$§ 8.2.1
Spike, $2\alpha + \beta \leq 1$	$\mathrm{poly}(n)$ [14]	$\exp(n)$ [13]	$O(1)$§ 8.3.1	$O(n)$§ 8.2.2
Spike, $2\alpha + \beta > 1$	$\exp(n)$ [14]	$\exp(n)$ [13]	$O(1)$§ 8.3.1	$O(n)$§ 8.2.2

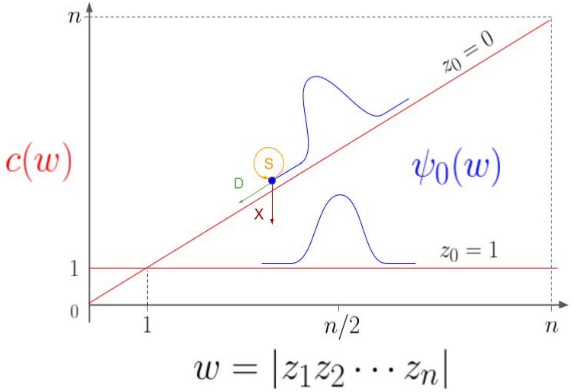
Table 1: Performance of the four algorithms, summarized. For the two instances studied, we distinguish different parameter regimes. For the Bush instance, the performance of QAO depends on the choice of mixer B_λ (see Eq. 22). For Spike, the QAO performance depends on spike parameters α and β. We see that bang-bang control algorithms outperform their (quantum and classical) annealing-based counterparts for these instances. Sources for existing results are cited, and the new contributions are referenced by the relevant sections.

Bang-bang v. quasistatic: classical

Claim: BBSA running pure gradient descent on the Bush finds the minimum efficiently.

Proof idea: Show a poly runtime by analyzing the discrete-time Markov chain.

Proof sketch:



Moves	*Probability*
Stay (S)	(n-w)/(n+1)
Descend (D)	w/(n+1)
Die (X)	1/(n+1)

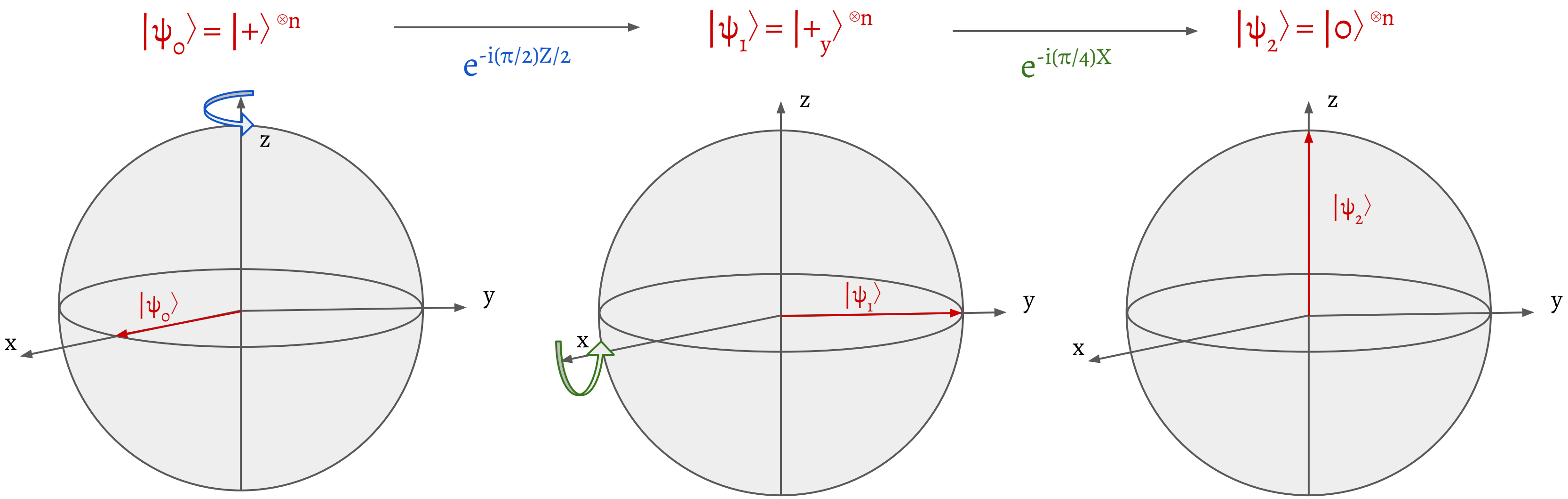
- Expected time to stay at w (under survival): $E[m_w \mid \bar{X}] = \Sigma_t \Pr(S^{t-1}D)$
- Expected number of moves (under survival): $E[m \mid \bar{X}] \leq \Sigma_w E(m_w \mid \bar{X}) = O(n\log(n))$
- Probability of reaching the minimum: $\Pr(\text{success}) = \Pr(\bar{X})^{E[m \mid \bar{X}]} = \Omega(1/n)$

Bang-bang v. quasistatic: quantum

Claim: A depth-2 QAOA circuit minimizes the Spike instance, for any $(\alpha,\beta) \in [0,1)\times[0,\infty]$.

Proof idea: We show that the wavefunction does not "see" the spike, only the ramp.

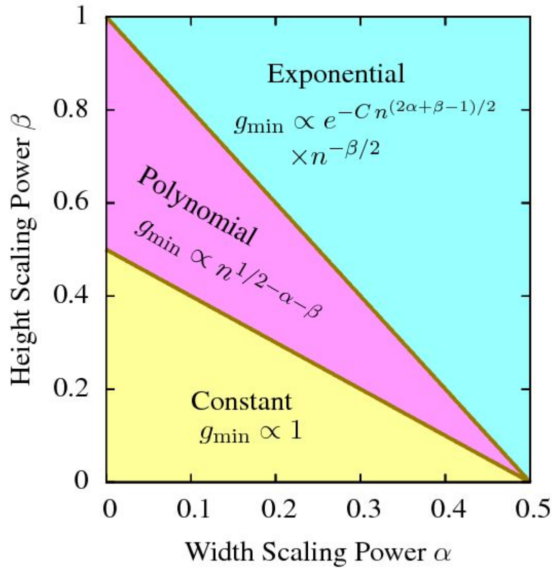
Under a simple ramp, the Hamiltonian is one-local, $H(u) = u\boldsymbol{\Sigma}_i Z_i/2 - (1-u)\boldsymbol{\Sigma}_i X_i$



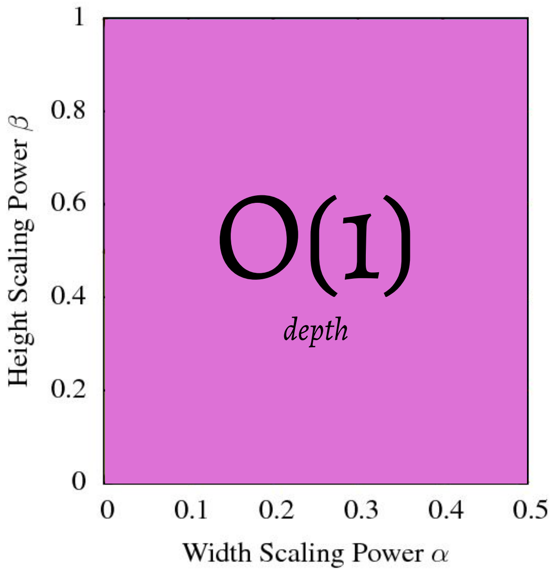
Bang-bang v. quasistatic: quantum

Claim: A depth-2 QAOA circuit minimizes the Spike instance, for any $(\alpha,\beta) \in [0,1)\times[0,\infty]$.

QAO



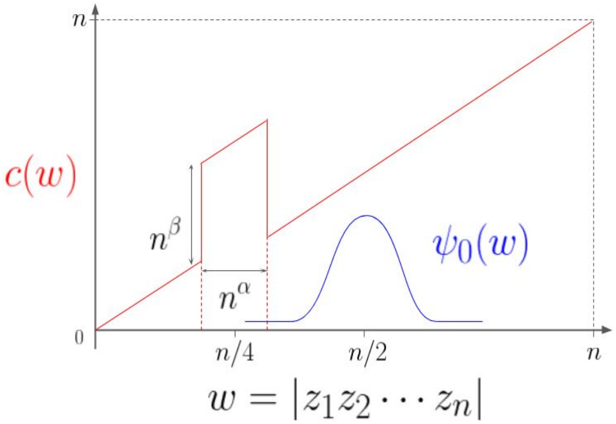
QAOA



Bang-bang v. quasistatic: quantum

Claim: A depth-2 QAOA circuit minimizes the Spike instance, for any $(\alpha,\beta) \in [0,1)\times[0,\infty]$.

Proof idea: We show that the wavefunction does not "see" the spike, only the ramp.



Lemma 1. *Let $c(w)$ be a cost function on Hamming weights, and let $p \in [0,1]$. Suppose $c(w) = r(w) + s(w)$, where r, s are two functions satisfying the following:*

1. $\min_w c(w) = \min_w r(w)$.
2. *There exist angles β, γ such that QAOA1 with schedule (β, γ) minimizes $r(w)$ with probability at least p.*
3. *If the initial state is $|\psi_0\rangle = \sum_w A_w |w\rangle$, then $s(w)$ overlaps weakly with $|\psi_0\rangle$ in the sense that*

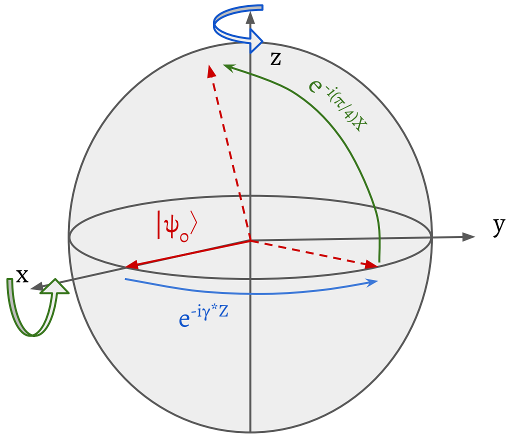
$$\sum_{w=1}^{n} 4|A_w|^2 \sin^2\left(\frac{\gamma s(w)}{2}\right) \equiv q \leq o(p)$$

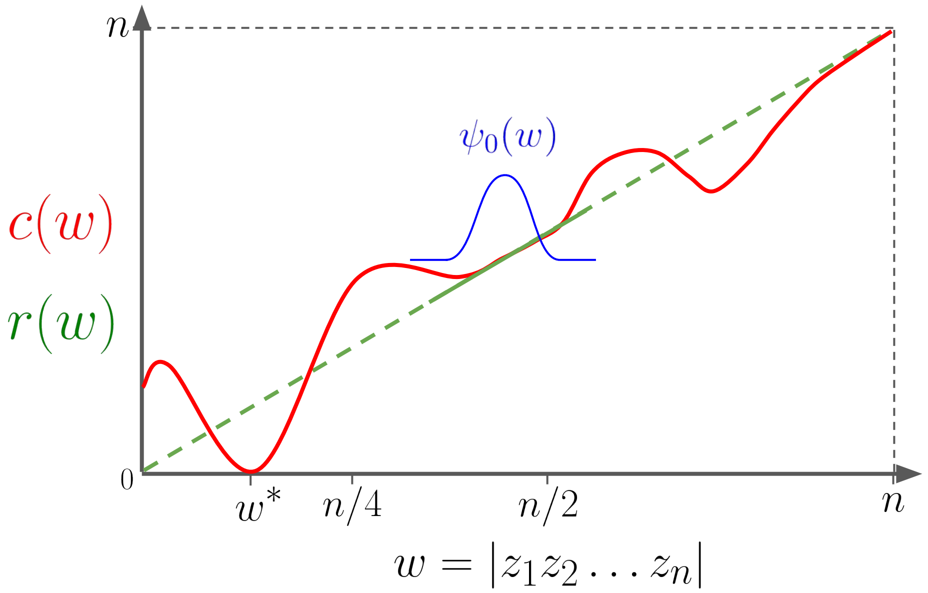
Then, QAOA1 with schedule (β, γ) minimizes $c(w)$ with probability at least $p - o(p)$.

QAOA on general symmetric instances

If a general Hamming-symmetric cost function is sufficiently "ramp-like" around w ~ n/2, we can try QAOA1 just like we did for the Spike instance. We need:

- Weak overlap,
- Slope at least 1/poly(n) around n/2

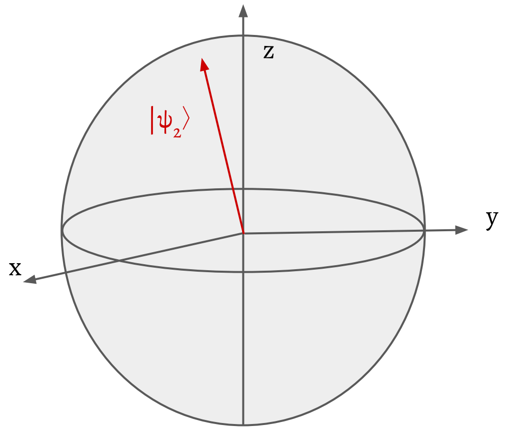


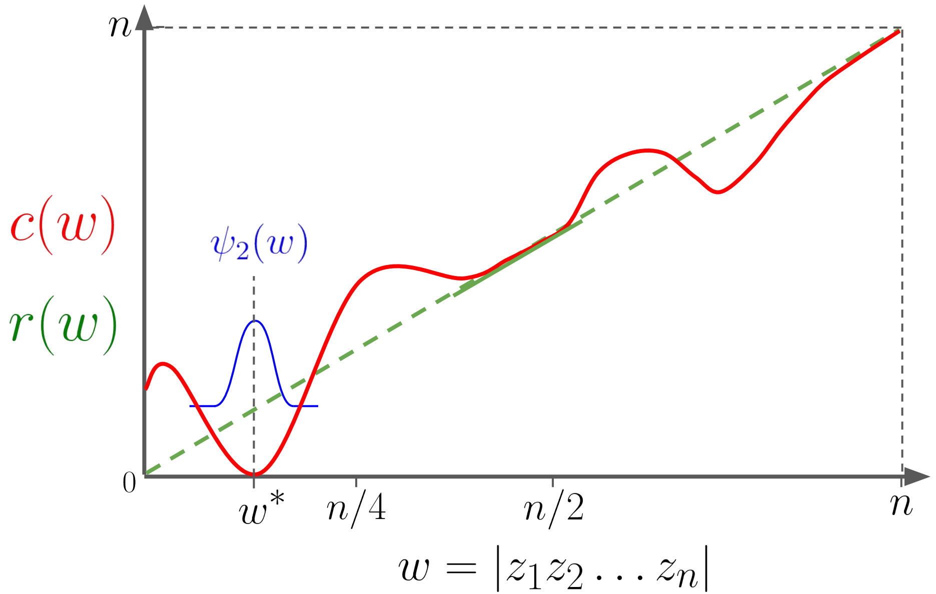


QAOA on general symmetric instances

If a general Hamming-symmetric cost function is sufficiently "ramp-like" around w ~ n/2, we can try QAOA1 just like we did for the Spike instance. We need:

- Weak overlap,
- Slope at least 1/poly(n) around n/2





Caveats to optimal control

While PMP itself is very generally applicable, the conclusion that linearly controlled optimal trajectories are bang-bang has certain caveats:

1. Singular intervals: The optimal value of the control is determined by the derivative of the control Hamiltonian w.r.t. the control:

$$u^*(t) = \begin{cases} 1, & \text{if } \partial_u \mathcal{H}(t) < 0 \\ 0, & \text{if } \partial_u \mathcal{H}(t) > 0 \end{cases}$$

   Singular (time) intervals are those in which the above derivative vanishes. Here, the optimal control remains indeterminate.
2. Infinite switches (aka *Fuller phenomenon*): The optimal bang-bang trajectory has an infinite number of switches, which renders the control infeasible. Seen in the optimal control of (analog) Grover search.

Thank you!