

Vliv symetrických vzorků na výsledky softwaru pro nalezení roviny symetrie

Technická zpráva

František Srb, Ivana Kolingerová

Vliv symetrických vzorků na výsledky softwaru pro nalezení roviny symetrie

Technická zpráva

František Srb, Ivana Kolíngerová

Abstract

Schopnost najít osu symetrie pro různorodá data přináší výhody pro další práci s těmito daty. Proto je dobré mít rychlý a spolehlivý software, které nám tyto osy najde. V této technické zprávě se zaměříme na jeden takovýto software a změříme kvalitu a přesnost nalezení os symetrie na předem připravených datech. Tyto výsledky zhodnotíme a pokusíme se upřesnit jak si software poradí s různými daty, která splňují různé podmínky.

Tato práce byla podpořena Grantovou agenturou České republiky, projekt GAČR 21-08009K Zobecněné symetrie a ekvivalence geometrických dat.

Kopie zprávy jsou dostupné na
<https://www.kiv.zcu.cz/cs/Research/Technical-reports.html>
nebo na žádost poslanou na následující adresu:

Západočeská univerzita v Plzni
Katedra informatiky a výpočetní techniky
Univerzitní 8
30614 Plzeň
Česká republika

Tímto by chtěli autoři poděkovat doc. RNDr. Michalu Bizzarimu, Ph.D. za jeho cenné odborné rady k problematice, které se tato technická zpráva věnuje.

Obsah

1	Úvod	5
2	Metodika	5
2.1	Metody	5
2.1.1	Metoda delší strany	5
2.1.2	Metoda náhodného doplnění	6
2.1.3	Metoda poměru	6
2.2	Funkce	6
3	Měření	6
3.1	Parametry	6
3.1.1	Měřené funkce	7
3.2	Výsledky měření	8
3.2.1	Metoda náhodného doplnění	9
3.2.2	Metoda větší strany (poměr 0.5)	12
3.2.3	Metoda větší strany (poměr 0.6)	15
3.2.4	Metoda větší strany (poměr 0.7)	18
3.2.5	Metoda větší strany (poměr 0.8)	21
3.2.6	Metoda větší strany (poměr 0.9)	24
3.2.7	Metoda větší strany (poměr 1)	27
3.2.8	Metoda poměru (vzorků 100)	30
3.2.9	Metoda poměru (vzorků 200)	33
3.2.10	Metoda poměru (vzorků 400)	36
4	Závěr	39

1 Úvod

V rámci této práce bude zkoumán vliv způsobu vzorkování na schopnost detekce osy symetrie v softwaru Ing. Lukáše Hrudý [1]. Pro dosažení tohoto cíle budou vybrány specifické funkce, které jsou symetrické podle jediné osy. Tyto funkce následně navzorkujeme podle určitých pravidel a parametrů a následně je předáme jako vstup softwaru. Výsledky nakonec porovnáme a vyvodíme závěr.

2 Metodika

V měření budou zohledněny variabilní hodnoty několika parametrů, konkrétně:

- poměr vzorků s obrazem ku všem vzorkům,
- metoda vzorkování a její případné parametry,
- počet vzorků,
- funkce a její transformace.

2.1 Metody

Použijeme tři různé metody na získání vzorků. U každé z nich předem víme, kolik vzorků budeme chtít vzorkovat.

2.1.1 Metoda delší strany

V datech označena jako *Larger_Side_Method*. U této metody máme dva parametry:

- *symmetry_ratio* neboli poměr symetrie – poměr vzorků, jejichž symetrický obraz je také součástí vzorků, ku všem vzorkům,
- *larger_side_ratio* neboli větší strana – zbylé vzorky jsou navzorkovány tak, aby poměr počtu vzorků z původní funkce bez jejího obrazu k celkovému počtu vzorků odpovídal tomuto poměru.

Jinými slovy, nejdříve vezmeme *symmetry_ratio* všech vzorků a navzorkujeme je tak, aby polovina z nich byla na původní funkci a druhá polovina odpovídala jejím obrazům. Následně zbylé vzorky rozdělíme tak, aby jich bylo *larger_side_ratio* na původní funkci a zbytek na obraze.

Příklad: Máme 100 vzorků, poměr symetrie 0.2 a větší stranu 0.8. Znamená to, že 20 vzorků bude symetrických (10 na původní funkci a 10 jejich obrazů

na obrazu funkce). Následně nám zbývá 80 vzorků, které rozdělíme tak, že na funkci bude 64 vzorků a na obrazu funkce 16.

2.1.2 Metoda náhodného doplnění

V datech označená jako *Random_Method*. Obdobně jako u předchozí metody nejprve navzorkujeme symetrické vzorky podle parametru *symmetry_ratio*. Poté vzorkujeme zbylé vzorky zcela náhodně z obou stran.

2.1.3 Metoda poměru

V datech označená jako *Ratio_Method*. Tato metoda může připomínat metodu větší strany, avšak zde se nejprve vzorkuje větší strana podle *larger_side_ratio* a až pak se doplní symetrické vzorky podle *symmetry_ratio*. Nakonec se dovzorkují chybějící vzorky na obrazu funkce.

2.2 Funkce

Funkce jsou popsány rovnicí $f(x, y^2) = 0$, kde $f(x, y)$ je libovolný polynom s aspoň jednou lichou mocninou proměnné x . Tyto funkce jsou symetrické podle osy x a neměly by být symetrické podle osy y .

Pro měření si vybereme několik funkcí, během měření budeme transformovat vzorky pomocí náhodné translace a rotace, což nám zaručí variaci symetrické osy.

3 Měření

V této sekci se nachází parametry měření a jejich výsledky.

3.1 Parametry

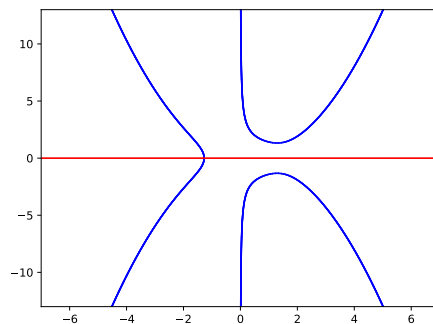
Testy byly spuštěny s těmito parametry:

- Metoda větší strany – s hodnotami parametru větší strany $[0.5, 0.6, 0.7, 0.8, 0.9, 1]$ a Metoda náhodného doplnění
 - Poměr symetrie – 0 až 0.68 po krocích velikosti 0.04 (celkem 18 hodnot)
 - Počet vzorků – 1 až 9.5 (krok 0.5) vzorků za jednotku šířky bounding boxu funkce (celkem 18 hodnot)
- Metoda symetrie měla následující hodnoty parametrů:
 - počet vzorků – $[100, 200, 400]$

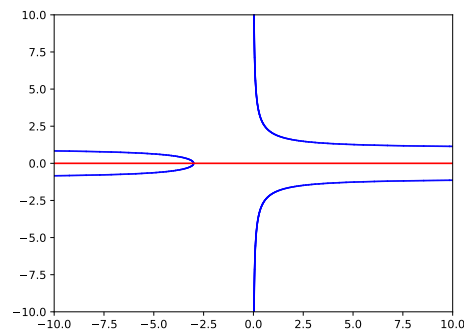
- větší strana – 0.5 až 0.98 po krocích velikosti 0.02 (celkem 25 hodnot)
- Poměr symetrie – 0 až 0.72 po krocích velikosti 0.03 (celkem 25 hodnot)

3.1.1 Měřené funkce

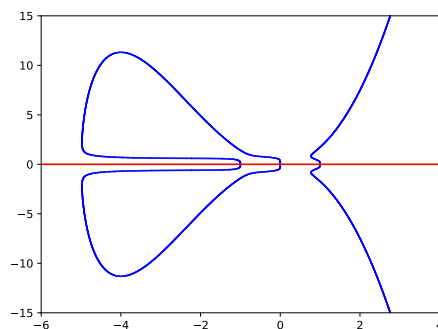
Seznam měřených funkcí (modře), jejich osy symetrie (červeně) a jejich bounding boxy (odpovídají prostoru grafu):



Obrázek 1: $f_1(x, y^2) = 0$, kde $f_1(x, y) = -x^5 + x^4 + 3xy - 6$



Obrázek 2: $f_2(x, y^2) = 0$, kde $f_2(x, y) = xy - x - 3$



Obrázek 3: $f_3(x, y^2) = 0$, kde $f_3(x, y) = x^5y^2 + 5x^4y^2 - 2y^3 + x^3 - x$

3.2 Výsledky měření

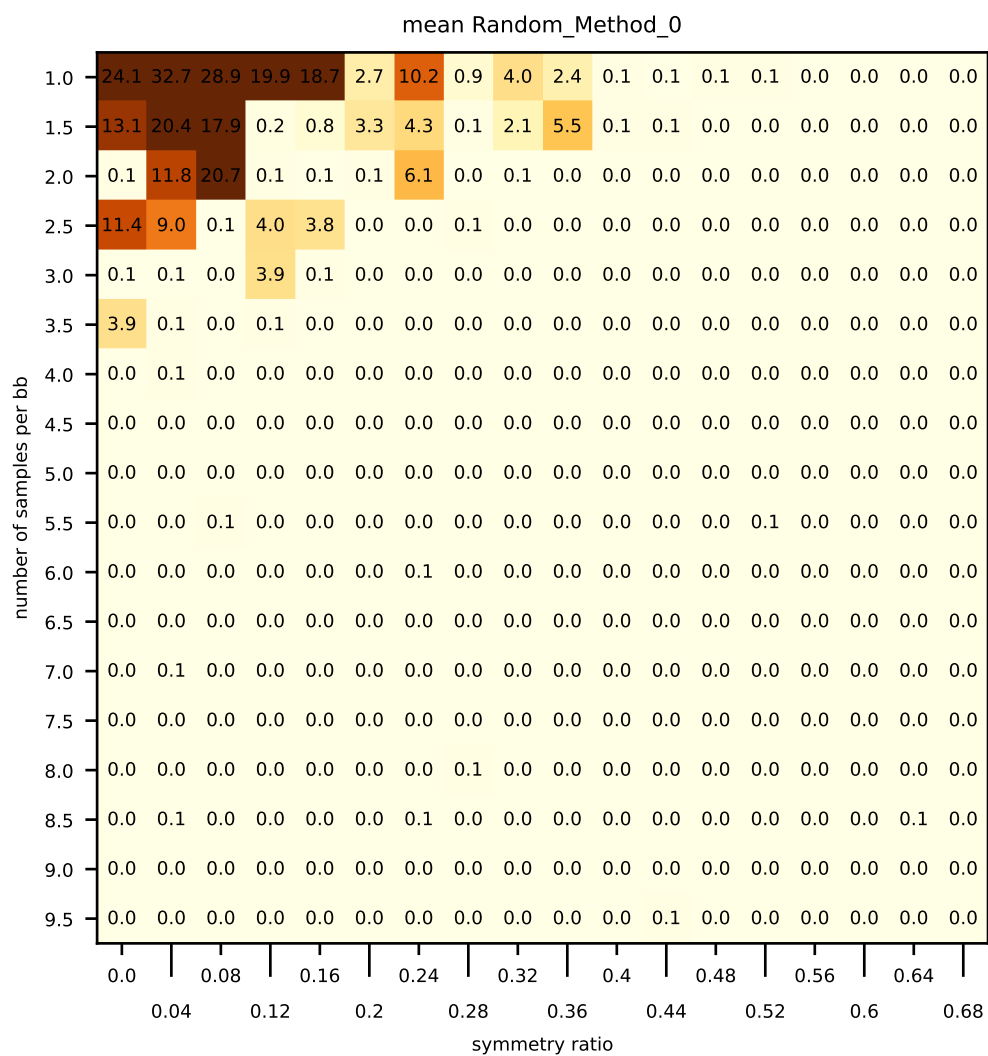
Níže uvedené tabulky prezentují seskupení informací týkajících se provedených měření. Název tabulky obsahuje metodu měření a také, zda se jedná o průměrnou hodnotu (*mean*) či o směrodatnou odchylku (*std*) naměřených hodnot. Buňka v tabulce pak znázorňuje průměrnou hodnotu či směrodatnou odchylku **rozdílů stupňů** výsledku softwaru vůči správnému řešení.

3.2.1 Metoda náhodného doplnění

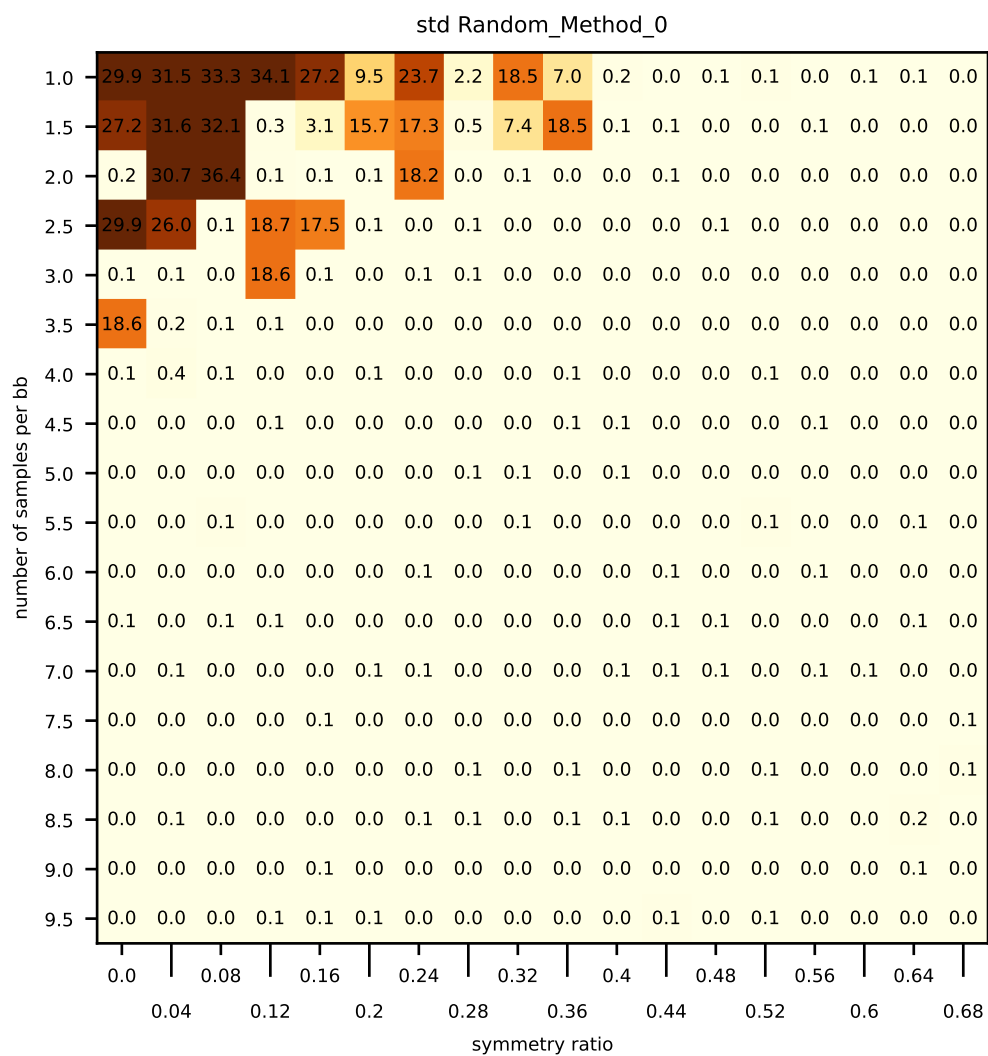
Metoda vzorkování byla popsána v sekci 2.1.2. Na obrázku 4 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 5 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

Na těchto datech vidíme, že software je odolný a pro náhodné vzorkování nachází správnou osu symetrie pro většinu kombinací parametrů (buňky s hodnotami 0.0 či 0.1). V levém horním rohu tabulky vidíme, že pro tyto kombinace parametrů software vrací poměrně špatné (průměrný rozdíl) a celkem náhodné (směrodatná odchylka) výsledky. Vidíme též, že pro počet vzorků pro hodnoty 2.5 a 3 se dramaticky zlepší výsledek softwaru. Můžeme tedy usoudit, že měření s parametrem počtu vzorků s hodnotami menší než 3 trpí podvzorkováním.

Podvzorkování však nemá negativní vliv na výsledek právě tehdy, když máme část vzorků navzorkovanou symetricky. Například pro hodnotu parametru pro symetrické vzorkování 0.2 a větší dostáváme až na pár výjimek dobré výsledky. Pro hodnotu 0.4 a větší dokonce nejsou v našich měření ani výjimky.



Obrázek 4: Průměrná hodnota *Metoda náhodného doplnění*



Obrázek 5: Směrodatná odchylka *Metoda náhodného doplnění*

3.2.2 Metoda větší strany (poměr 0.5)

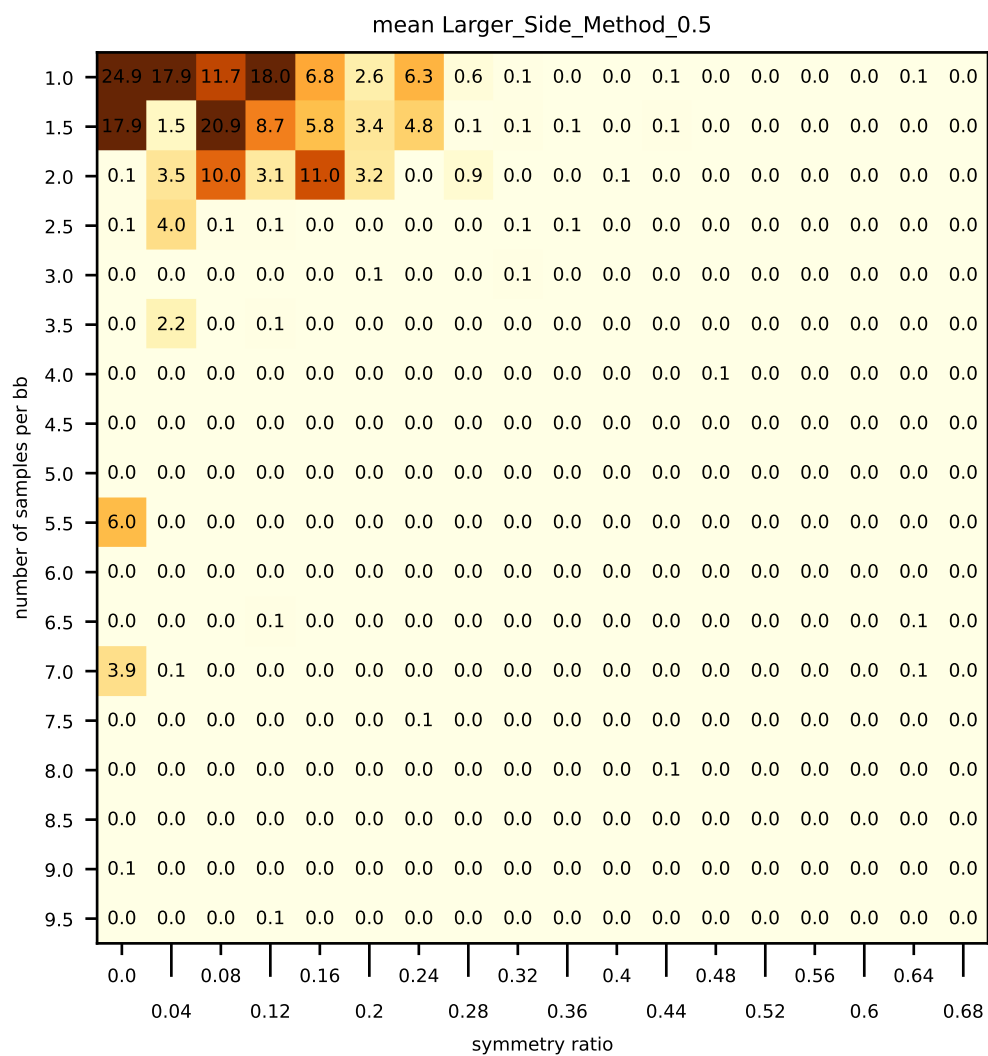
Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 6 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 7 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

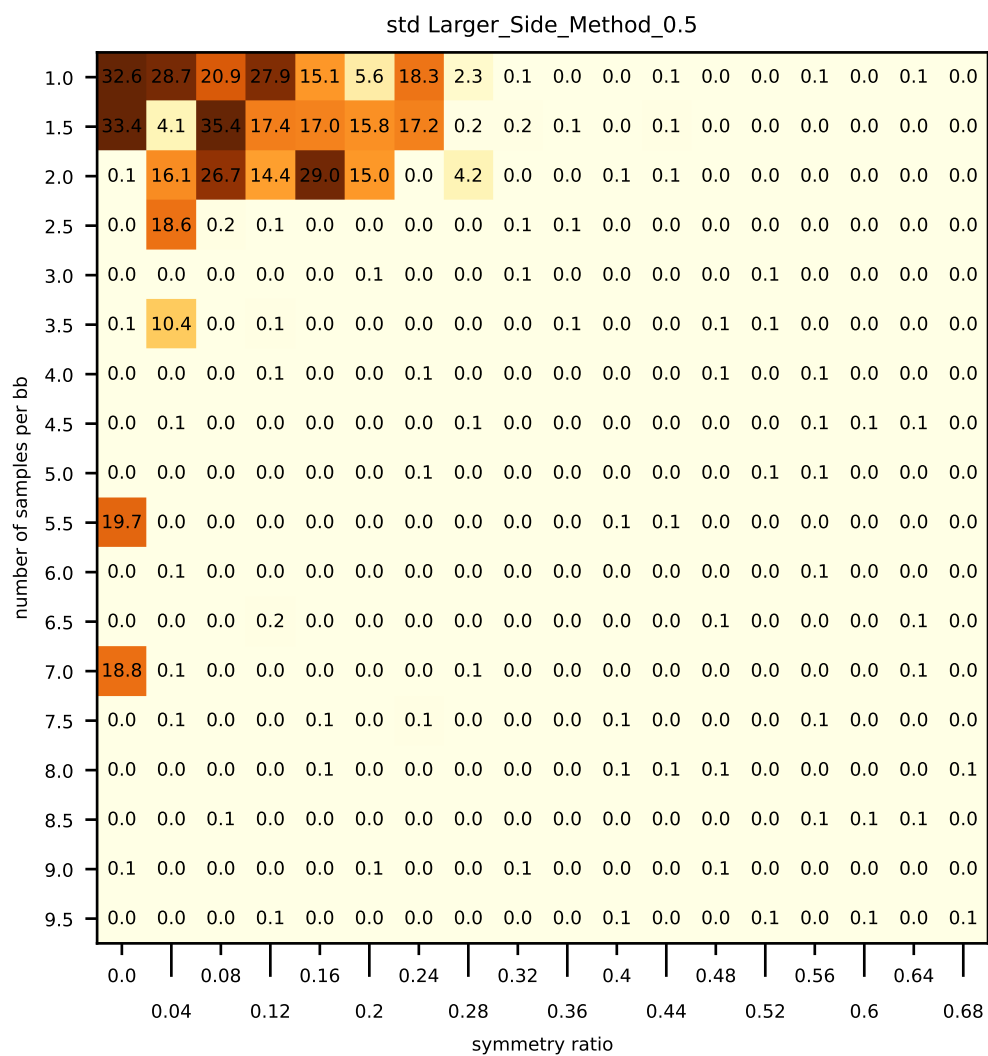
- $n \cdot \text{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot (1 - \text{symmetry_ratio})$ jsou ostatní vzorky, které jsou rovnoměrně (půl na půl) rozděleny mezi funkci i její obraz.

Na těchto datech vidíme, že software si lehce s takovými vzorky poradí (buňky s hodnotami 0.0 či 0.1). V levém horním rohu tabulky vidíme, že pro tyto kombinace parametrů software poskytuje poměrně špatné (průměrný rozdíl) a celkem náhodné (směrodatná odchylka) výsledky. Vidíme též, že pro počet vzorků pro hodnotu 2.5 se dramaticky zlepšil výsledek softwaru. Můžeme tedy usoudit, že měření s parametrem počtu vzorkování s hodnotami menší než 2.5 opět trpí podvzorkováním.

I zde vidíme, že částečná symetrie vzorků (hodnoty 0.2 či 0.28 a větší) nám pomůže získat dobré výsledky i za předpokladu podvzorkování.



Obrázek 6: Průměrná hodnota *Metoda větší strany (poměr 0.5)*



Obrázek 7: Směrodatná odchylka *Metoda větší strany* (poměr 0.5)

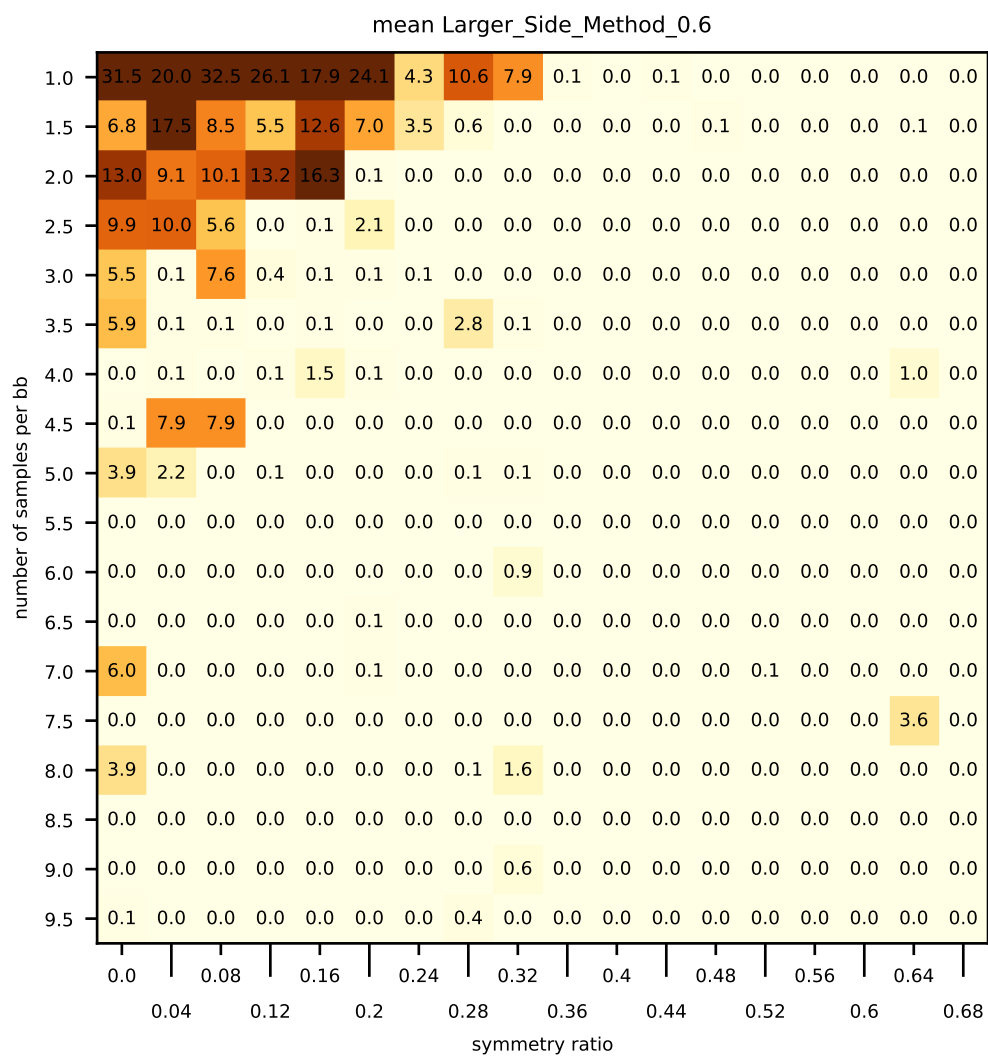
3.2.3 Metoda větší strany (poměr 0.6)

Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 8 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 9 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

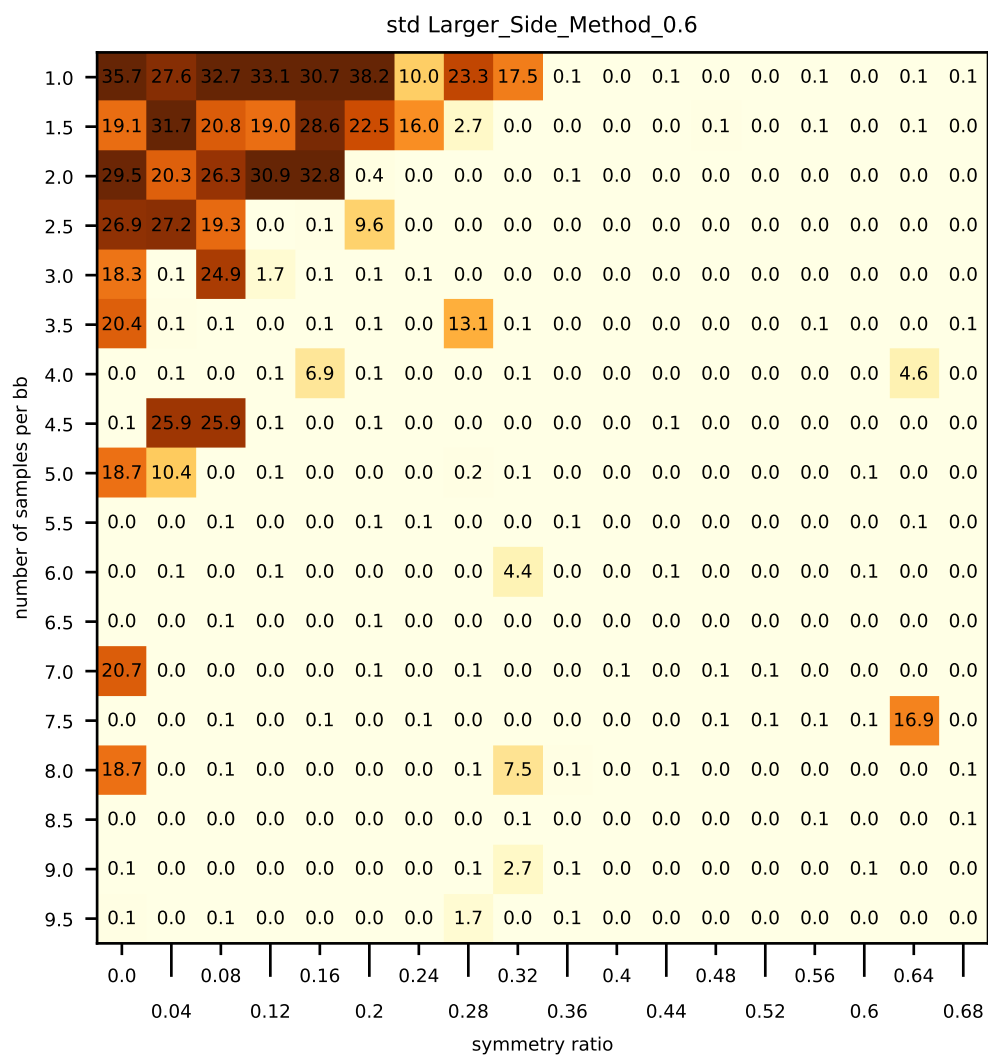
Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

- $n \cdot \textit{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot 0.2 \cdot (1 - \textit{symmetry_ratio})$ jsou vzorky, které se všechny nacházejí pouze na jedné straně osy symetrie
- $n \cdot 0.8 \cdot (1 - \textit{symmetry_ratio})$ jsou ostatní vzorky, které jsou rovnoměrně (půl na půl) rozděleny mezi obě strany osy symetrie.

Zde už pozorujeme lehké, ale stále minimální zhoršení oproti předchozím metodám. Především je důležité zmínit kombinaci parametrů pro symetrii 0.0 - 0.08 a počet vzorků 2.5 a 3, kde nyní pozorujeme častější chybu a také pro počet vzorků 1.0 nyní potřebujeme už symetrii 0.36, abychom vynahradili podvzorkování.



Obrázek 8: Průměrná hodnota *Metoda větší strany (poměr 0.6)*



Obrázek 9: Směrodatná odchylka *Metoda větší strany (poměr 0.6)*

3.2.4 Metoda větší strany (poměr 0.7)

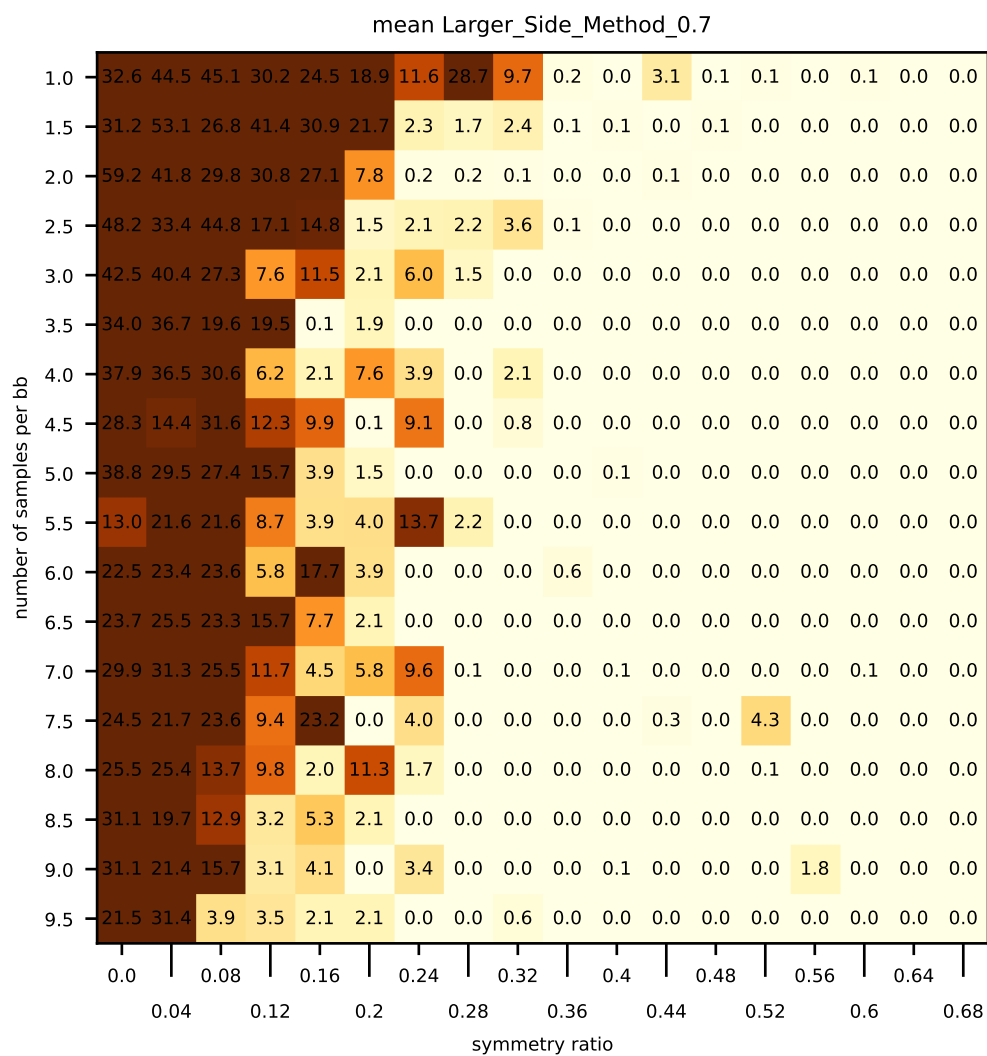
Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 10 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 11 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

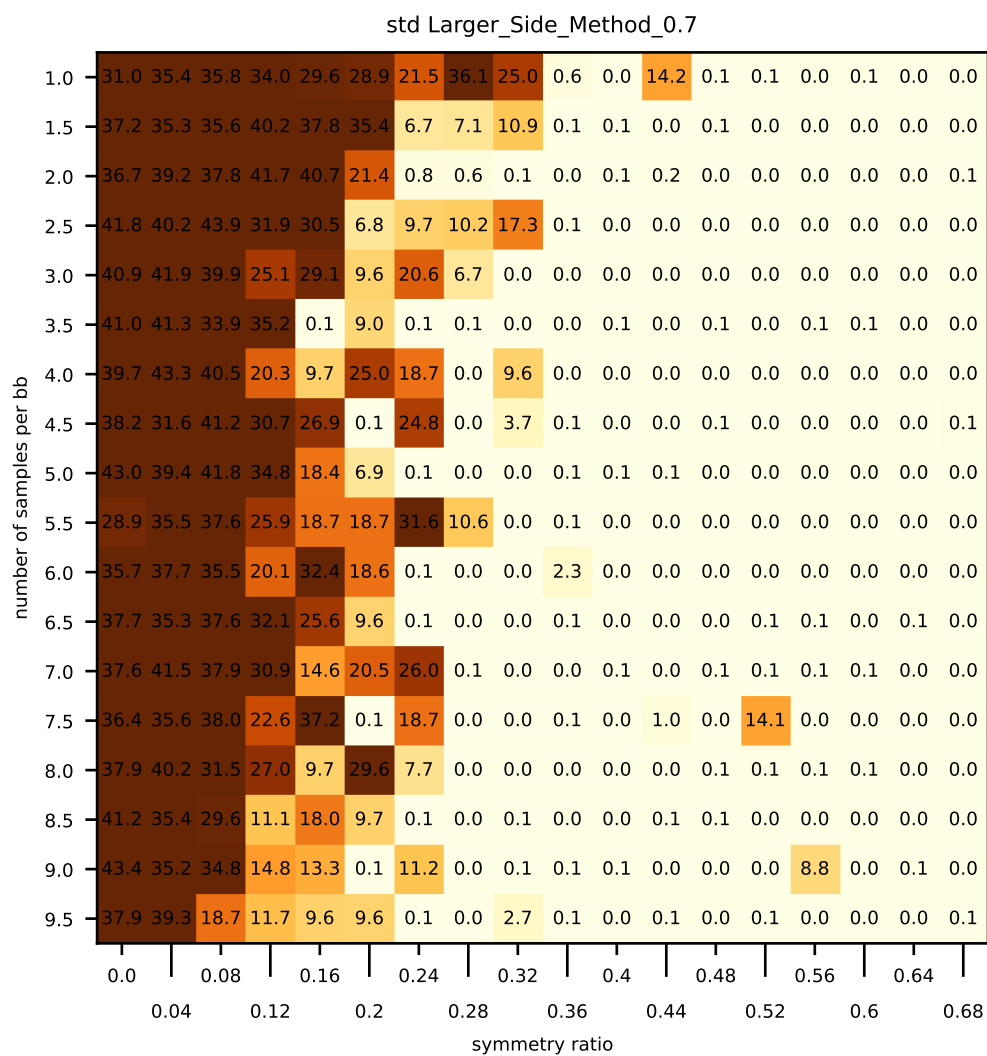
- $n \cdot \text{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot 0.4 \cdot (1 - \text{symmetry_ratio})$ jsou vzorky, které se všechny nacházejí pouze na jedné straně osy symetrie
- $n \cdot 0.6 \cdot (1 - \text{symmetry_ratio})$ jsou ostatní vzorky, které jsou rovnoměrně (půl na půl) rozděleny mezi obě strany osy symetrie.

Zde už vidíme větší chyby i u kombinací parametrů, které předtím netrpěly problémy s podvzorkováním. U těchto dat pozorujeme, že pokud má jedna strana symetrie méně jak třetinu všech vzorků, program nenajde osu symetrie. Pokusíme si vypořádat tuto hypotézu i u pozdějších metod.

Důležité je poznamenat, že se v našem měření nejdříve naměří symetrické vzorky a až na zbylé se použije zmíněná metoda. Například pro parametr poměru symetrie 0.2 a 1000 vzorků se nejdříve navzorkuje 100 vzorků na jedné straně osy a dalších 100 jejich obrazů a až pak se na zbylých 800 vzorků aplikuje metoda vzorkování.



Obrázek 10: Průměrná hodnota *Metoda větší strany* (poměr 0.7)



Obrázek 11: Směrodatná odchylka *Metoda větší strany (poměr 0.7)*

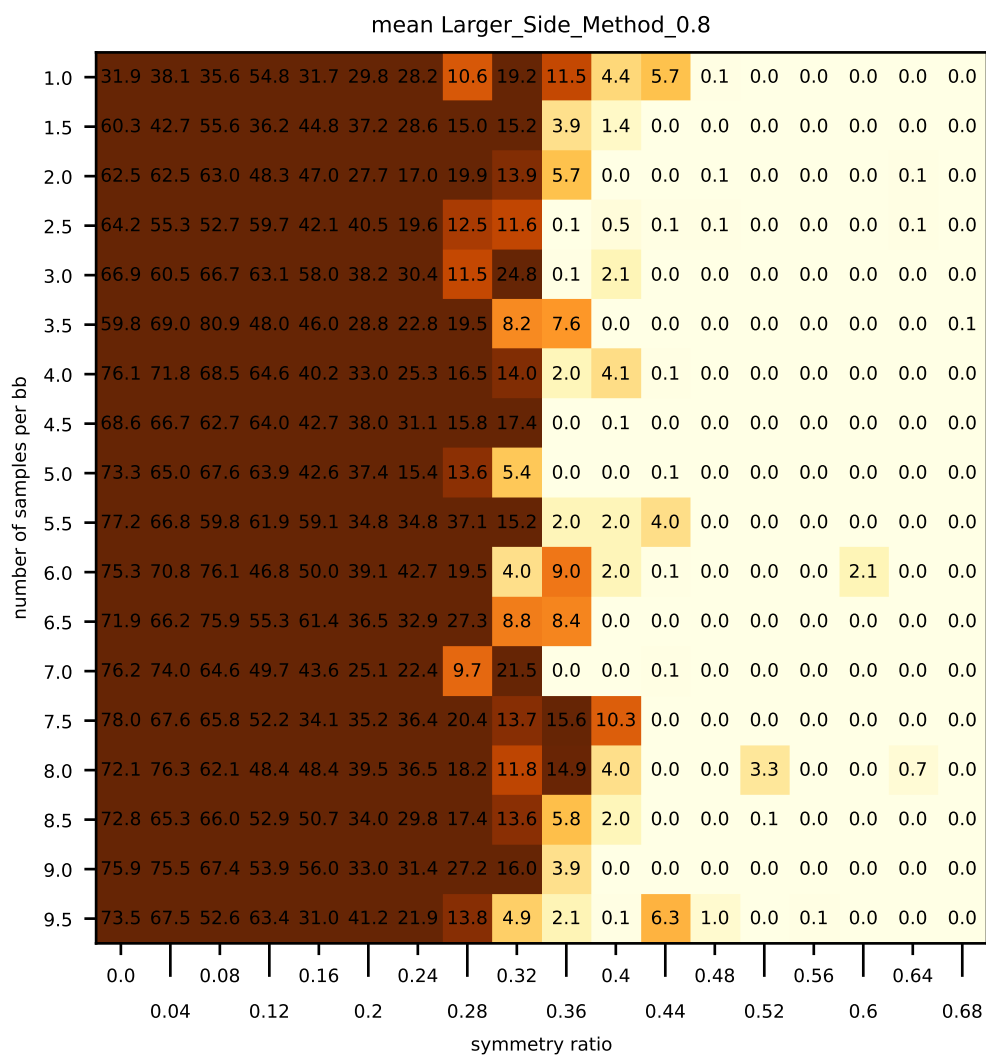
3.2.5 Metoda větší strany (poměr 0.8)

Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 12 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 13 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

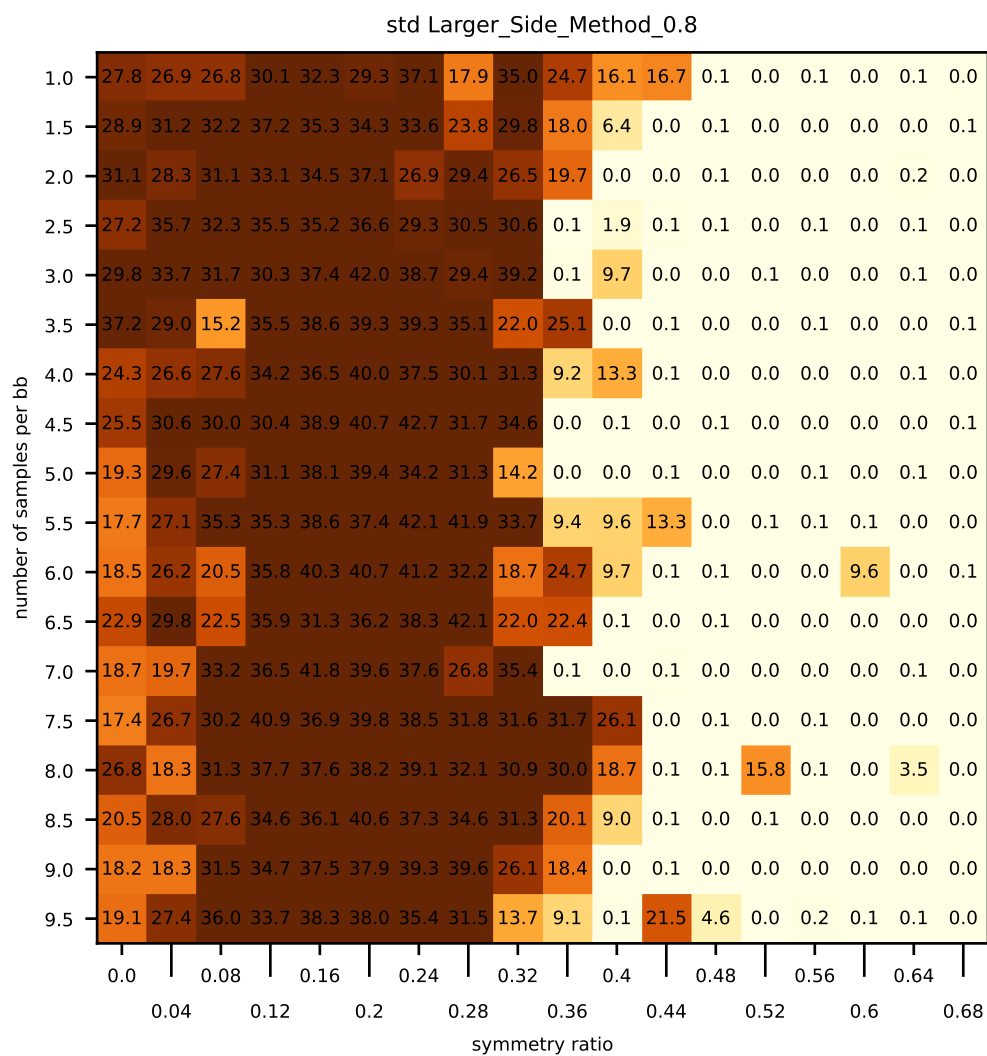
Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

- $n \cdot \text{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot 0.6 \cdot (1 - \text{symmetry_ratio})$ jsou vzorky, které se všechny nacházejí pouze na jedné straně osy symetrie
- $n \cdot 0.4 \cdot (1 - \text{symmetry_ratio})$ jsou ostatní vzorky, které jsou rovnoměrně (půl na půl) rozděleny mezi obě strany osy symetrie.

Zde se předchozí chyba rozšířila až do hodnot parametru poměru symetrie 0.32-0.4. U těchto dat opět pozorujeme, že pokud má jedna strana symetrie méně jak zhruba třetinu všech vzorků, program nenajde osu symetrie.



Obrázek 12: Průměrná hodnota *Metoda větší strany (poměr 0.8)*



Obrázek 13: Směrodatná odchylka *Metoda větší strany (poměr 0.8)*

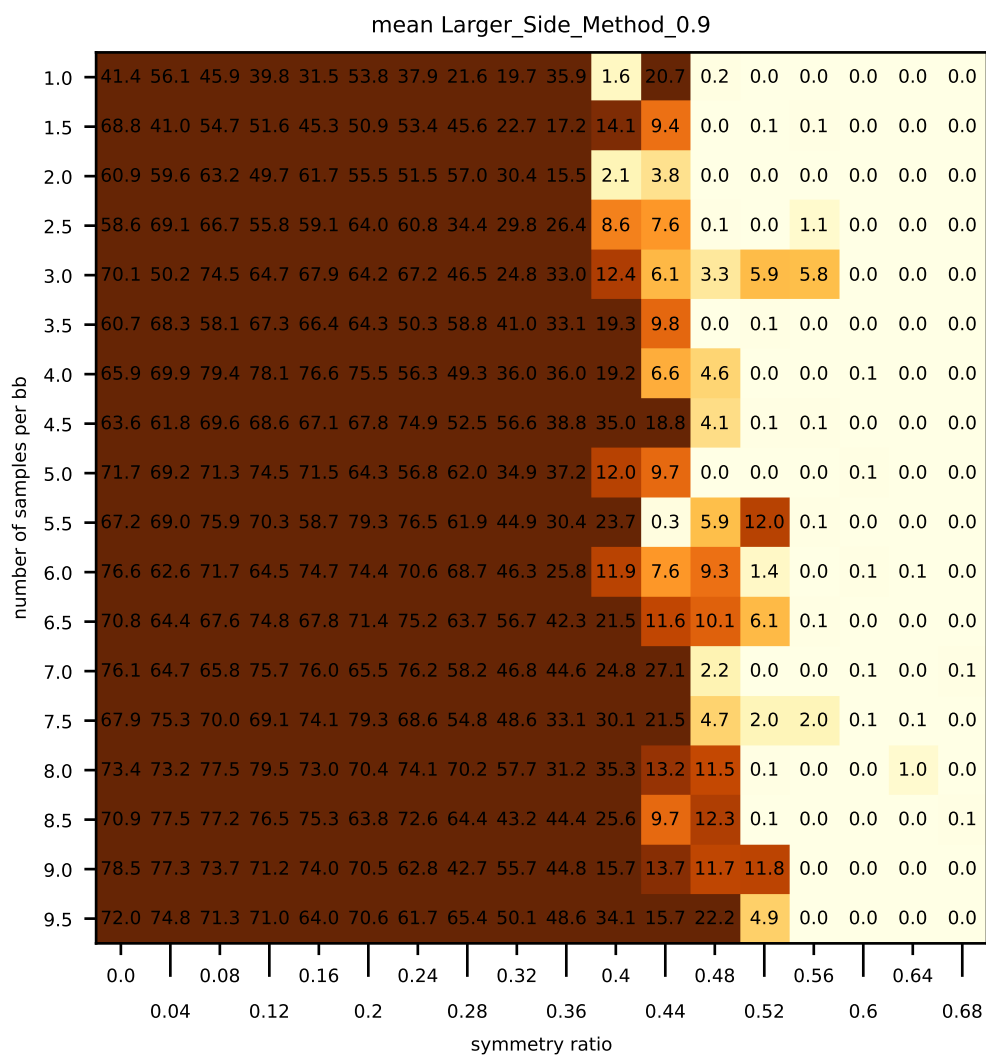
3.2.6 Metoda větší strany (poměr 0.9)

Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 14 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 15 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

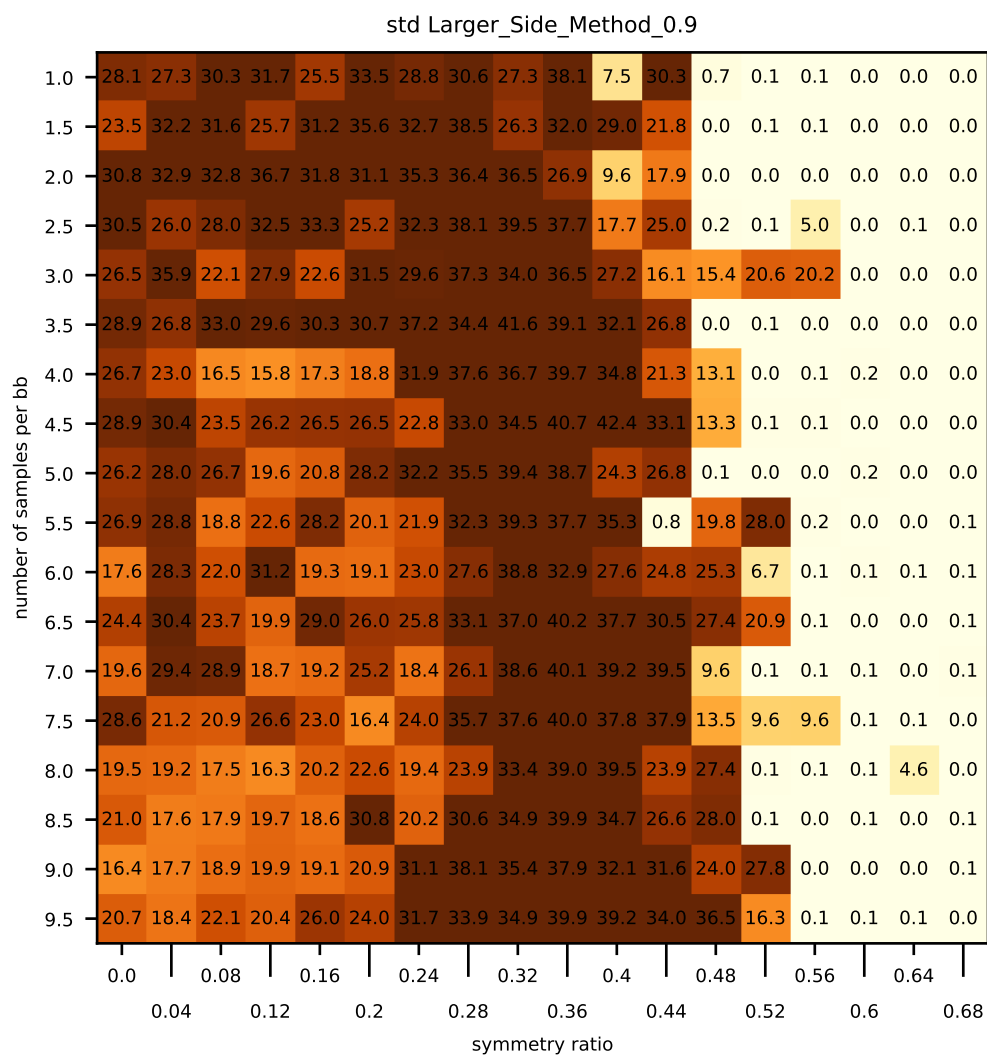
Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

- $n \cdot \text{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot 0.8 \cdot (1 - \text{symmetry_ratio})$ jsou vzorky, které se všechny nacházejí pouze na jedné straně osy symetrie
- $n \cdot 0.2 \cdot (1 - \text{symmetry_ratio})$ jsou ostatní vzorky, které jsou rovnoměrně (půl na půl) rozděleny mezi obě strany osy symetrie.

Zde se nám opět chyba rozšířila až do hodnot parametru poměru symetrie 0.44-0.52. I zde se nám opět ukázalo, že pokud má jedna strana symetrie méně jak zhruba třetinu všech vzorků, program nenajde osu symetrie.



Obrázek 14: Průměrná hodnota *Metoda větší strany* (poměr 0.9)



Obrázek 15: Směrodatná odchylka *Metoda větší strany (poměr 0.9)*

3.2.7 Metoda větší strany (poměr 1)

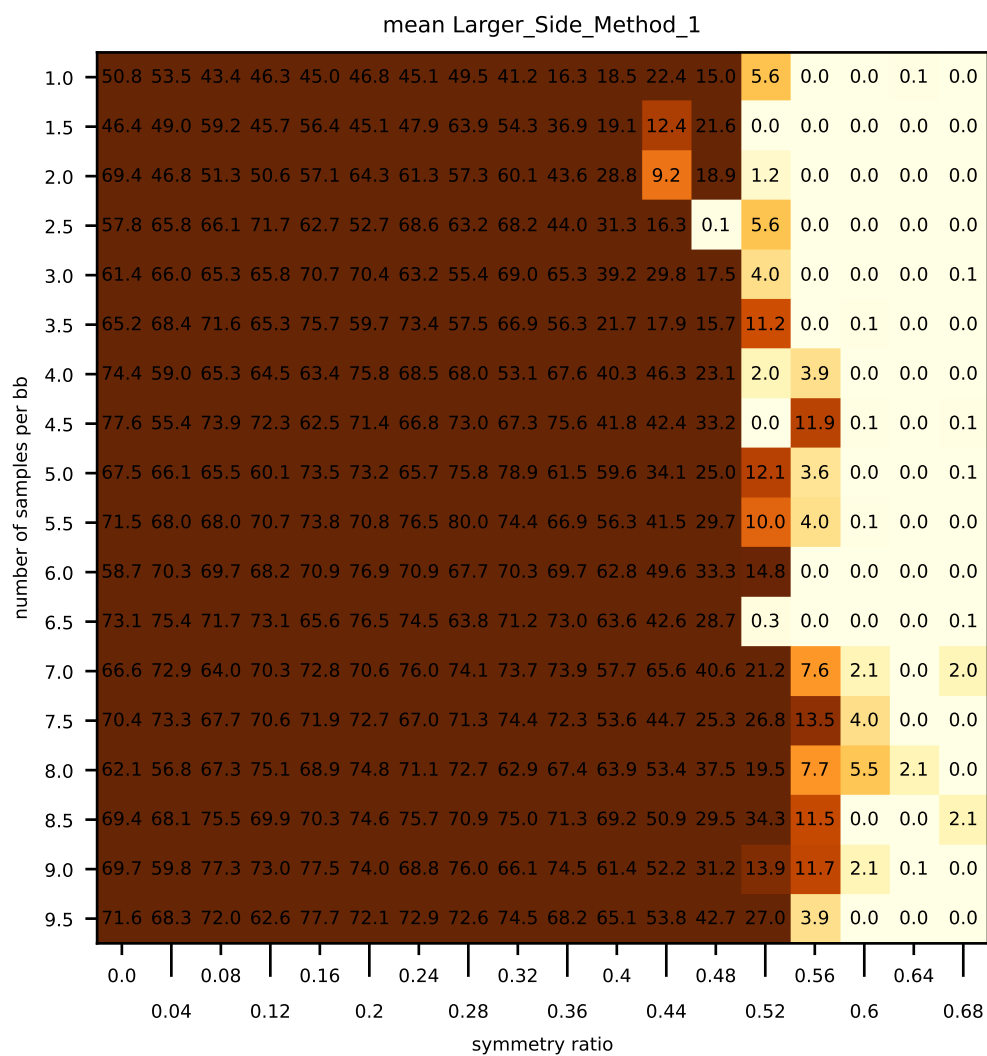
Metoda vzorkování byla popsána v sekci 2.1.1. Na obrázku 16 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 17 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení.

Data pro tuto metodu vypadají následovně (n je celkový počet vzorků pro danou instanci):

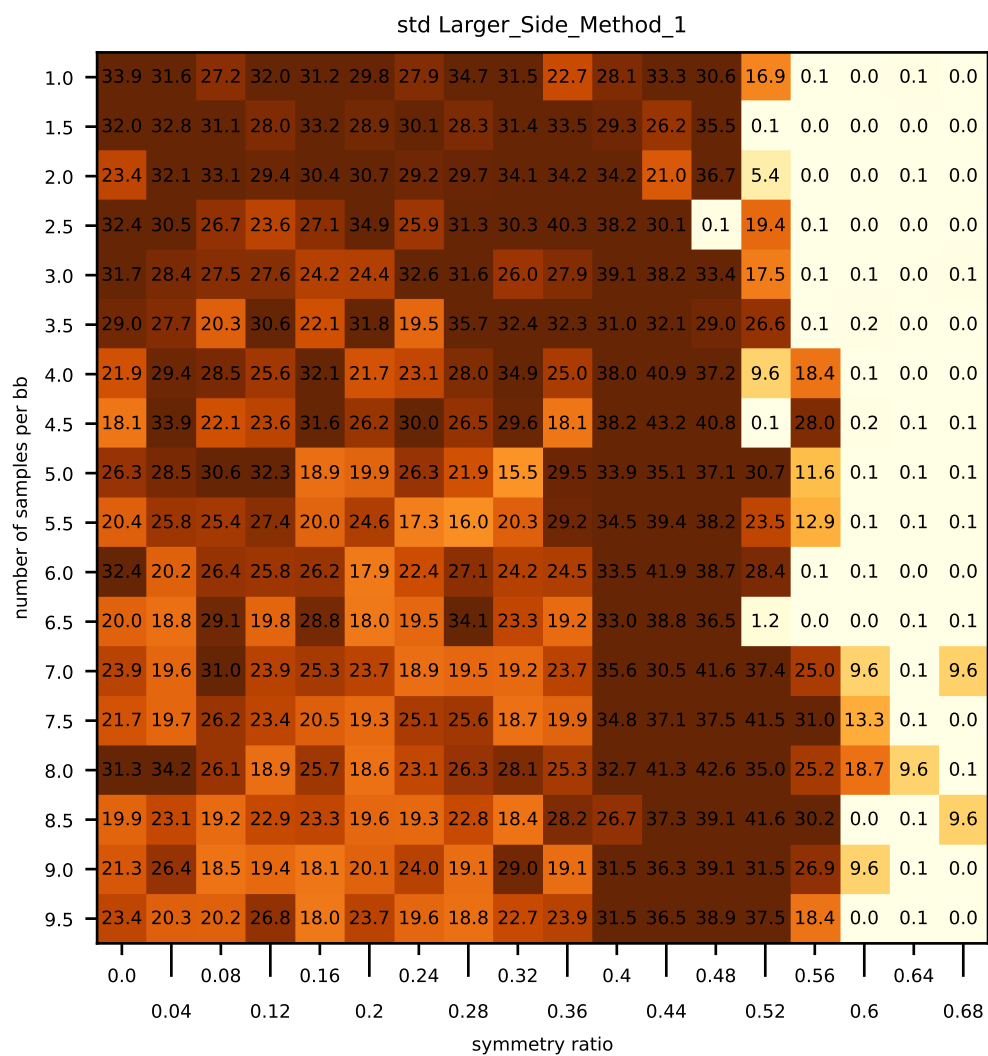
- $n \cdot \text{symmetry_ratio}$ jsou symetricky naměřené vzorky, kde pro každý vzorek existuje jiný vzorek, který je jeho obrazem.
- $n \cdot (1 - \text{symmetry_ratio})$ jsou vzorky, které se všechny nachází pouze na jedné straně symetrie.

Z měření pozorujeme, že data, která obsahují polovinu či méně symetrických vzorků, nejsou vhodná pro nalezení osy symetrie. Pozorujeme, že software si těžko poradí s daty, kde více jak polovina vzorků se nachází na jedné straně symetrie.

Toto měření také bylo důvodem pro vytvoření další metody testů, konkrétně metody poměru.



Obrázek 16: Průměrná hodnota *Metoda větší strany (poměr 1)*



Obrázek 17: Směrodatná odchylka *Metoda větší strany (poměr 1)*

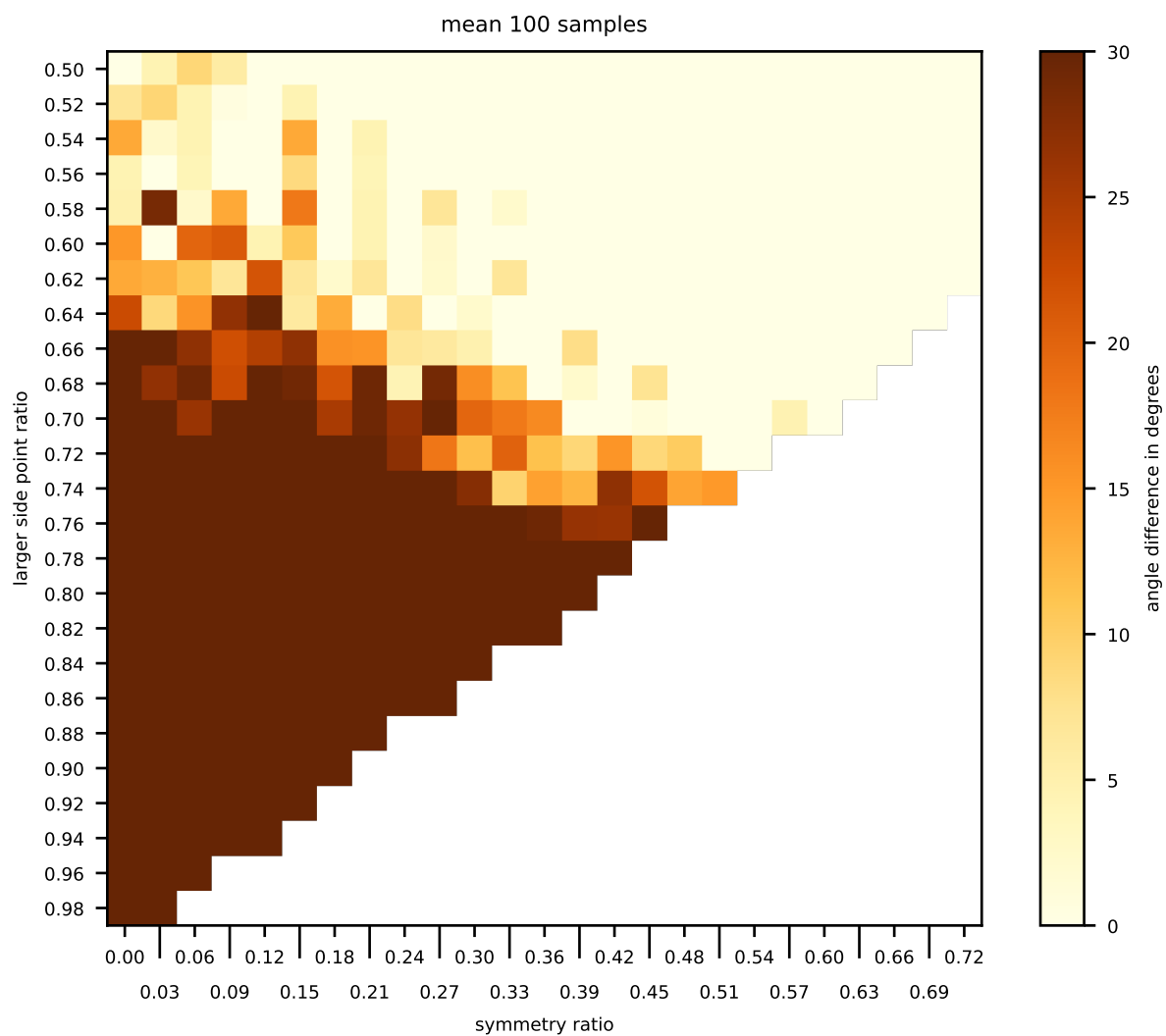
3.2.8 Metoda poměru (vzorků 100)

Metoda vzorkování byla popsána v sekci 2.1.3. Na obrázku 18 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 19 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení. Bílé hodnoty značí kombinace parametrů, pro které nelze vytvořit data splňující tyto parametry.

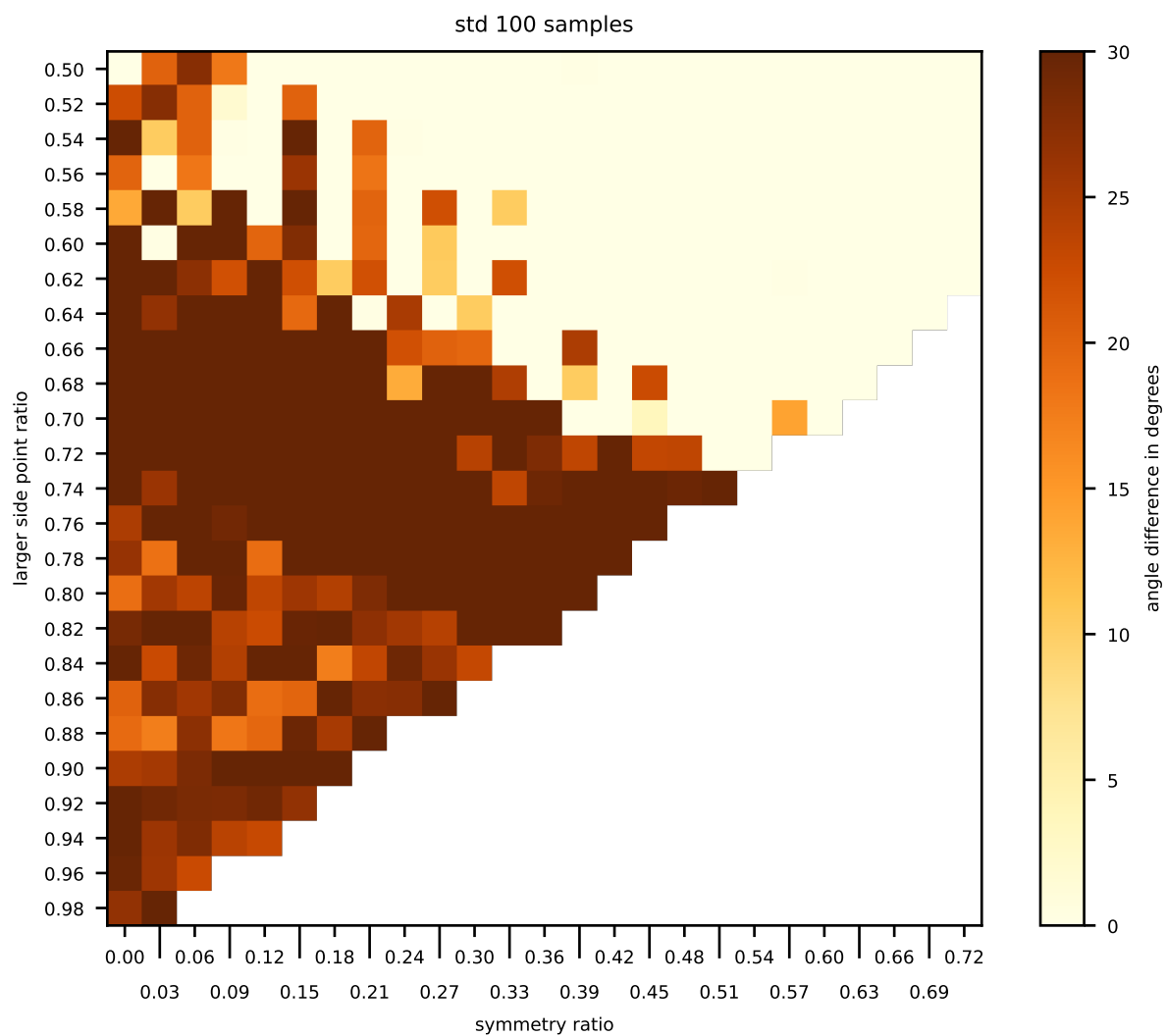
Data pro tuto metodu vypadají následovně (celkem 100 vzorků při každém měření):

- $100 \cdot larger_side_ratio$ je počet vzorků nacházející se na jedné straně osy symetrie,
- $100 \cdot symmetry_ratio$ je počet vzorků, které mají svůj obraz také ve vzorcích.

Pozorujeme, že software začne mít problémy v případě, kdy jedna strana obsahuje alespoň $\frac{2}{3}$ všech vzorků. S tímto si do jisté míry dokážou poradit symetricky vzorkovaná data. Konkrétně si software neporadil s daty, která obsahovala $\frac{3}{4}$ všech vzorků na jedné straně osy symetrie.



Obrázek 18: Průměrná hodnota *Metoda poměru* (vzorků 100)



Obrázek 19: Směrodatná odchylka *Metoda poměru* (vzorků 100)

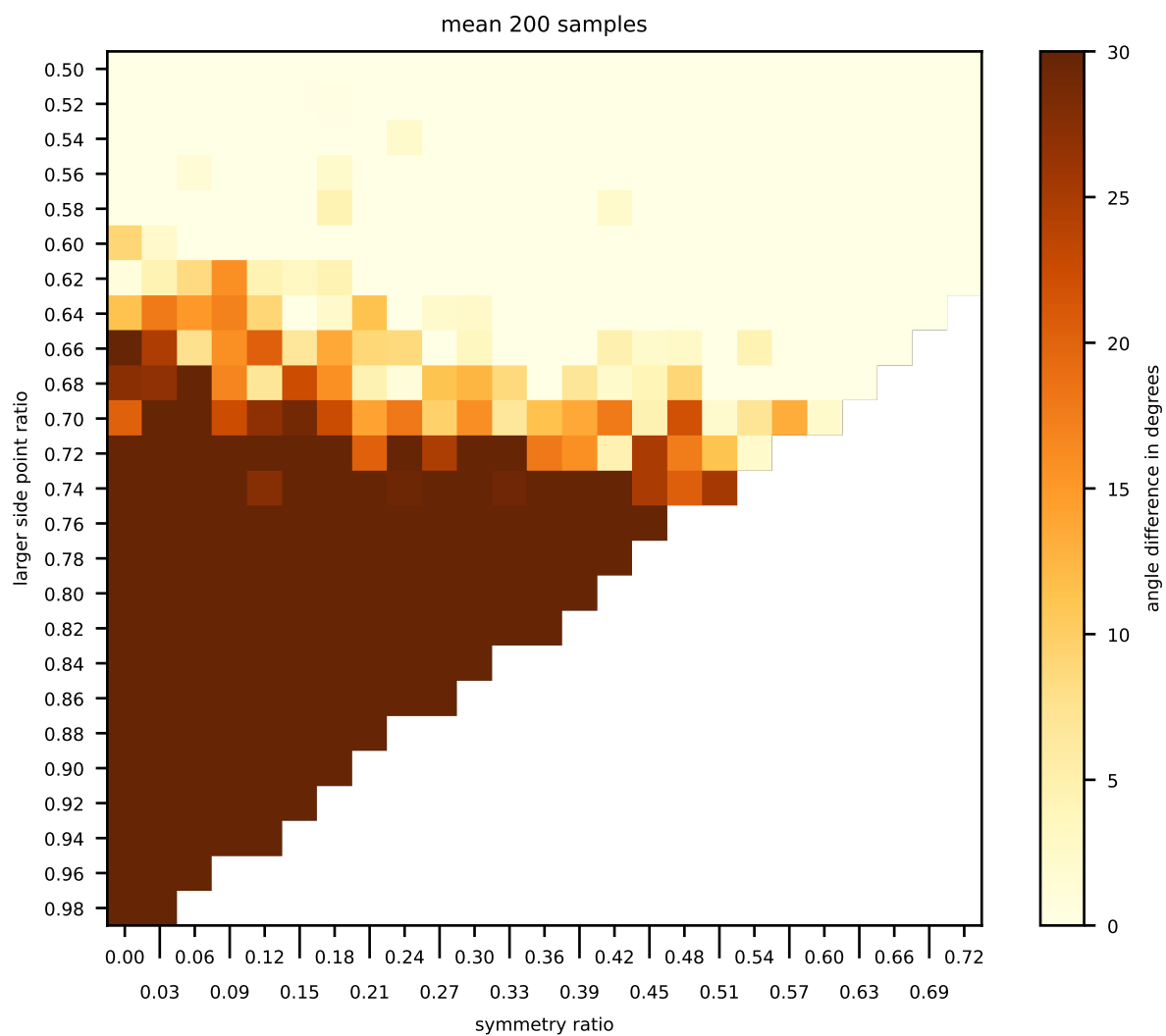
3.2.9 Metoda poměru (vzorků 200)

Metoda vzorkování byla popsána v sekci 2.1.3. Na obrázku 20 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 21 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení. Bílé hodnoty značí kombinace parametrů, pro které nelze vytvořit data splňující tyto parametry.

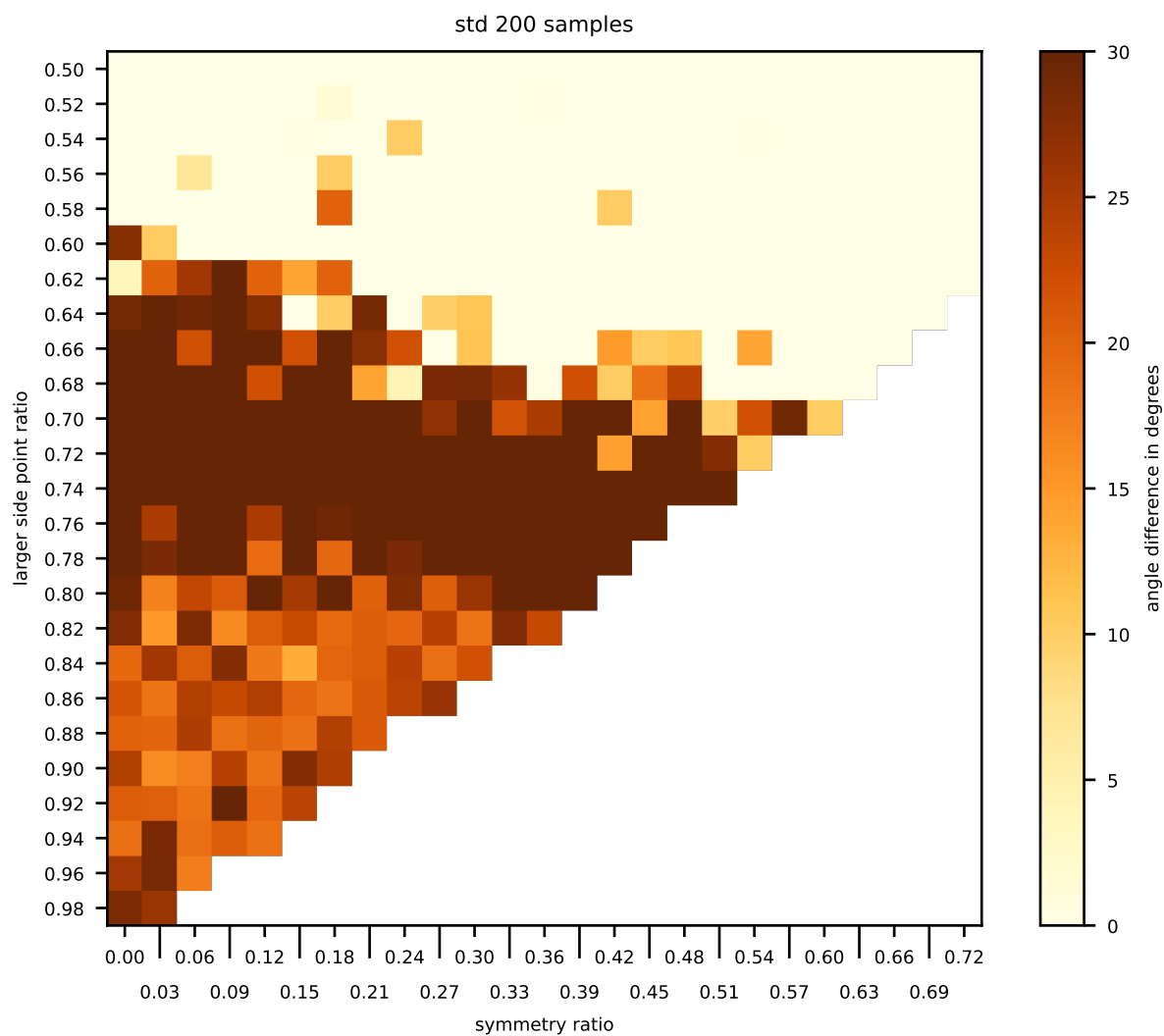
Data pro tuto metodu vypadají následovně (celkem 200 vzorků při každém měření):

- $200 \cdot larger_side_ratio$ je počet vzorků nacházející se na jedné straně osy symetrie,
- $200 \cdot symmetry_ratio$ je počet vzorků, které mají svůj obraz také ve vzorcích.

I zde pozorujeme, že software začne mít problémy v případě, kdy jedna strana obsahuje alespoň $\frac{2}{3}$ všech vzorků. S tímto si do jisté míry dokáže poradit symetricky vzorkovaná data. Konkrétně si software neporadil s daty, která obsahovala $\frac{3}{4}$ všech vzorků na jedné straně osy symetrie.



Obrázek 20: Průměrná hodnota *Metoda poměru (vzorků 200)*



Obrázek 21: Směrodatná odchylka *Metoda poměru* (vzorků 200)

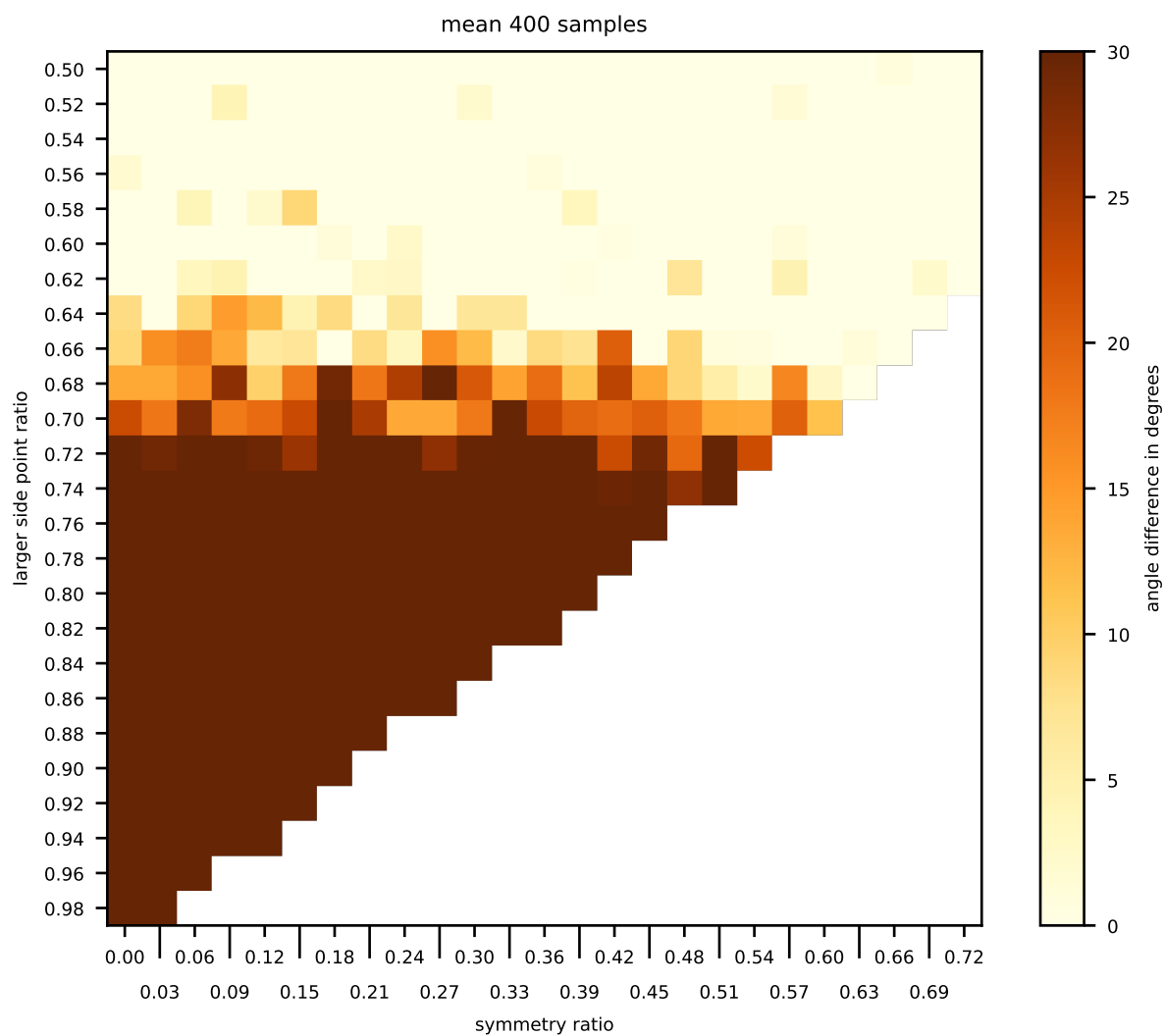
3.2.10 Metoda poměru (vzorků 400)

Metoda vzorkování byla popsána v sekci 2.1.3. Na obrázku 22 vidíme průměrné hodnoty výsledků měření pro různé parametry vzorkování a na obrázku 23 vidíme směrodatné odchylky těchto měření. Čím je barva políčka v obrázku tmavší, tím větší je rozdíl naměřených úhlů od správného řešení. Bílé hodnoty značí kombinace parametrů, pro které nelze vytvořit data splňující tyto parametry.

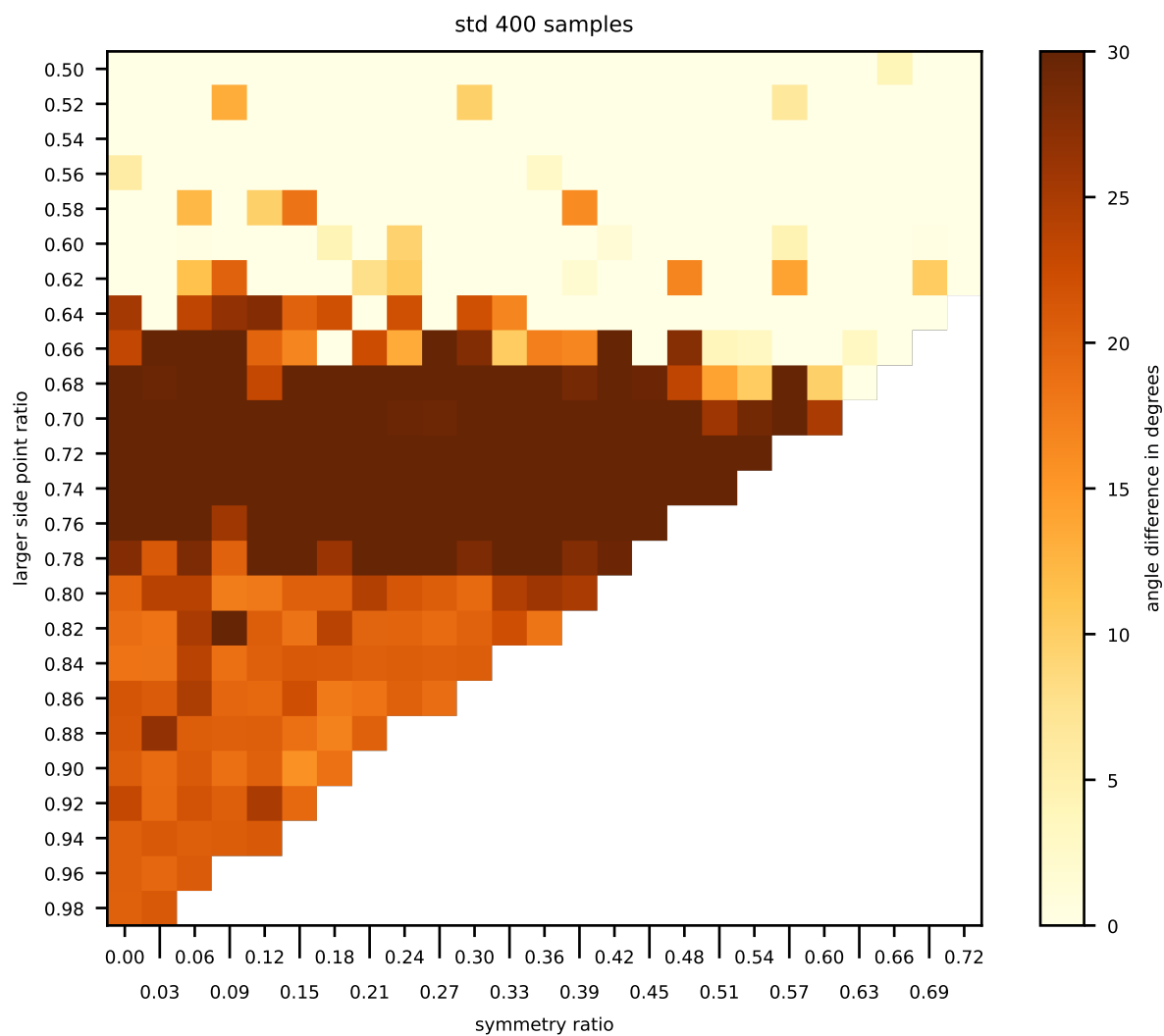
Data pro tuto metodu vypadají následovně (celkem 400 vzorků při každém měření):

- $400 \cdot larger_side_ratio$ je počet vzorků nacházející se na jedné straně osy symetrie,
- $400 \cdot symmetry_ratio$ je počet vzorků, které mají svůj obraz také ve vzorcích.

Novinkou u tohoto pozorování je fakt, že software začal mít problémy až u dat, kde 68% – 70% všech vzorků se nachází na jedné straně. V tomto případě symetrie vzorků pomohla velice málo či vůbec. Z tohoto pozorování můžeme usoudit, že symetrické vzorkování dokáže nahradit problémy s podvzorkováním, ale chtělo by to potvrdit dalšími měřeními.



Obrázek 22: Průměrná hodnota *Metoda poměru* (vzorků 400)



Obrázek 23: Směrodatná odchylka *Metoda poměru* (vzorků 400)

4 Závěr

Dohromady jsme spustili software na více než 50 000 různých vzorkování.

Z výsledků můžeme pozorovat, že symetrické vzorkování dokáže nahradit jinou chybu ve vzorkování, jako je například podvzorkování.

Z prvních měření jsme zjistili, jaká data dělají softwaru problém, a následně jsme potvrdili, že tento problém lze vyřešit zvětšením počtu vzorků či symetrickým vzorkováním.

Také jsme přišli na to, že software si nedokáže poradit s daty, která obsahují velké artefakty ve vzorkování. Konkrétně jsme pozorovali, že u vzorků, které se nacházejí z alespoň tří čtvrtin na jedné straně osy symetrie, nelze spolehlivě najít osu symetrie.

Z těchto dat jde tedy usoudit, že symetrie ve vzorkování má vliv na výstup softwaru a dokáže do jisté míry nahradit jiné chyby způsobu vzorkování.

Reference

- [1] Lukáš HRUDA. *Detekce symetrie v geometrických modelech [online]*. Doctoral theses, University of West Bohemia, Faculty of Applied Sciences Plzeň, 2023 [cit. 2024-06-18]. Supervisor: Prof. Dr. Ing. Ivana Kolingerová.