

Prof. Ing. Václav Hlaváč, CSc.

Doc. Ing. Miloš Sedláček, CSc.

# ZPRACOVÁNÍ SIGNÁLŮ A OBRAZŮ

2002  
Vydavatelství ČVUT

**Vydavatelství ČVUT upozorňuje autory na dodržování autorských práv.**

**Za jazykovou a věcnou správnost obsahu díla odpovídá autor. Text neprošel jazykovou ani redakční úpravou.**

© Václav Hlaváč, Miloš Sedláček, 2000  
ISBN 80-01-02114-9

# Předmluva

Skriptum Zpracování signálu a obrazu je určeno pro studenty stejnojmenného předmětu zavedeného do výuky oboru kybernetika a měření při přestavbě studia v letním semestru školního roku 1999/2000. Předmět seznamuje se základy číslicového zpracování signálů, a to jak jednorozměrných (obvykle průběhů v čase), tak i dvojrozměrných (často obrazů).

Pro elektrotechnického inženýra patří zpracování a analýza signálů mezi základní znalosti, které se od třicátých let dvacátého století využívají v radiotechnice, telekomunikacích, měření, regulaci a jinde. Části této disciplíny jsou řadu let vyučovány na elektrotechnické fakultě ČVUT v různých podobách a předmětech. Za tradiční lze považovat zpracování analogových signálů. Od sedesátých let získává díky číslicovým obvodům, počítačům a signálovým procesorům stále větší důležitost zpracování signálů diskrétních v čase (zejména signálů kvantovaných také v amplitudě, čili signálů číslicových).

V předmětu počítáme se studentovými znalostmi získanými v matematice (lineární algebře, analýze v komplexním oboru, numerické matematice, teorii lineárních integrálních transformací, a to zejména transformaci Fourierově), v teorii řízení a teorii obvodů (spojitý a diskrétní signál, deterministický a náhodný signál, průchod signálu lineárním systémem, popis signálu ve frekvenční oblasti, použití Fourierovy transformace a Z transformace). Zájemce o realizaci metod v signálových procesorech odkazujeme na navazující předměty.

Předmět se skládá ze dvou přibližně stejně obsáhlých částí.

První část (kap. 1 až kap. 5) napsaná M. Sedláčkem pokrývá nejpodstatnější teoretické části zpracování jednorozměrných signálů. Navazuje na předmět vyučovaný v letech 1991–1998 na katedře měření pod názvem Zpracování signálů v měřicí technice a následně Zpracování signálů. V části věnované zpracování stochastických signálů se vzhledem k důležitosti těchto signálů v měření a řízení věnuje pozornost částečně i analogovým metodám zpracování těchto signálů.

Podrobnější informace k teorii zpracování jednorozměrných signálů (včetně partií nepokrytých první částí skripta) lze získat např. v českých pracích [86, 76, 92, 91, 43] nebo v anglicky psaných knihách, např. v [68, 12, 39, 97, 40, 72]. Pro zájemce o hlubší proniknutí do teorie zpracování diskrétních signálů lze doporučit zejména monografi [68].

Druhá část skripta napsaná V.Hlaváčem se věnuje zpracování dvojrozměrných signálů, tj. obrazů. Postupy vysvětlené v první části skripta pro jednorozměrné signály se zde zobecňují pro obrazy. V celém skriptu se předpokládá, že signál není interpretován. Výhodou takového přístupu je, že postupy nejsou vázané na konkrétní aplikace. Zde popsané metody zpracování obrazu jsou využívané v počítačovém vidění, které bere v úvahu interpretaci obrazových dat a trojrozměrnou povahu obrazy reprezentovaného světa kolem nás. Čtenář se širšimu přístupu bude moci věnovat v navazujících předmětech (např. v předmětu Počítačové vidění a virtuální realita). Hlubší pozornost je věnována radiometrickým okolnostem nutným pro pochopení jevů ovlivňujících pořízení obrazu a šedotónové matematické morfologii. Obě téma jsou v české literatuře zpracována poprvé.

Další informace ke zpracování obrazů lze získat v česky a slovensky psaných knihách [32] (je již rozebraná a psána až příliš populární formou), nebo [44, 31, 82, 48, 43]. Z anglicky psaných knih zmiňme knihu [89] (jejíž kapitoly jsou psané V. Hlaváčem byly inspirací i pro toto skriptum), a [14, 34, 42, 77, 19, 21, 41].

Přejeme čtenáři, ať se mu předmět Zpracování signálu a obrazu dobře studuje.

Václav Hlaváč, Miloš Sedláček

## **Poděkování**

Autoři děkují doc. Ing. P. Sovkovi, CSc. z katedry teorie obvodů za pečlivou recenzi rukopisu. Dále děkuji svým kolegům z Centra strojového vnímání katedry kybernetiky (Ing. T. Pajdlovi, Dr. Ing. R. Šárovi, Dr. Ing. J. Matasovi, Ing. V. Smutnému a Ing. Tomášovi Svobodovi) a z katedry měření (Ing. R. Šmídovi, Dr. Ing. J. Holubovi a Ing. J. Blaškovi), se kterými zpracování signálů a zpracování obrazů zkoumají a učí. Ing. D. Hejtmanové, CSc. z katedry měření děkuji autoři za pečlivé přečtení výsledné verze rukopisu. V. Hlaváč děkuje dále prof. M. Šonkovi (Univ. of Iowa, USA) a R. Boyle (Univ. of Leeds, GB), protože při spolupráci na obou vydáních knihy [89] vznikal jeho názor na vykládanou látku.

## **Opravené chyby ve druhém dotisku**

První vydání tohoto skripta vyšlo v lednu 2000. První dotisk prvního vydání byl vytiskněn v únoru 2001 a bylo opraveno jen pár překlepů, a to uvnitř dosavadních stran.

Podklady pro dosud nejnovější druhý dotisk byly připraveny v červenci 2002. Kapitoly 1–6 týkající se zpracování signálu (M. Sedláček) nebyly změněny. V části o zpracování obrazu byly učiněny následující opravy:

*Kapitola 8, strany 130-133:* v textu i obrázcích byla řada oprav vycházející z chybné záměny ohniska a hlavního bodu při zavedení modelu čočky v geometrické optice.

*Kapitola 11, strany 160-161:* v příkladu 11.1 demonstrujícím separabilní filtr bylo opraveno pořadí násobení vektorů. Příklad byl mírně rozšířen o odvození, které usnadní pochopení, jak se uspoří výpočty.

*Kapitola 11, strana 164:* mírně opraven text uvozující Laplaceán  $\nabla^2$ .

*Kapitola 12, strana 183:* v definici transformace tref či miň bylo změněno nesprávné "... Za druhé nesmí ..." správným "... Za druhé musí ...".

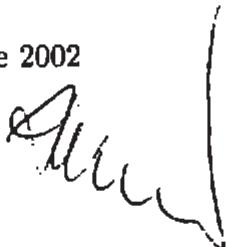
*Kapitola 12, strana 190:* legenda obr. 12.20, část (c), kde je nesprávné otevření nahrazeno správnou dilatací s vysvětlením, že jde o otevření výchozího obrázku.

*Kapitola 12, strana 193:* mírně upraven text zavádějící ztenčování.

Po vložení oprav se při automatickém formátování text na některých stranách posunul. Od prvního dotisku skripta se liší následující strany: ii, 130-135, 160-167, 183, 190, 193 a rejstřík na 214-220.

Za upozornění na chyby děkuji kolegům Vladimíru Smutnému, Ondřeji Drbohlavovi a studentovi Ivanovi Havlovi.

Václav Hlaváč, 15. července 2002



# Obsah

<b>1</b>	<b>Úvod části o zpracování signálů</b>	<b>1</b>
<b>2</b>	<b>Vzorkování signálu a rekonstrukce</b>	<b>4</b>
2.1	Vzorkovací věta . . . . .	4
2.2	Diracův impuls a jeho vlastnosti . . . . .	5
2.3	Odbození průběhu spektra vzorkovaného signálu . . . . .	7
2.4	Rekonstrukce spojitého signálu z posloupnosti vzorků . . . . .	8
2.4.1	Rekonstrukce pomocí ideální dolnofrekvenční propusti . . . . .	8
2.4.2	Rekonstrukce pomocí vzorkovače s pamětí (Lagrangeova interpolátoru řádu 0) . . . . .	10
2.4.3	Rekonstrukce pomocí lineárního interpolátoru . . . . .	11
2.4.4	Chyba rekonstrukce signálu . . . . .	12
2.5	Bloková struktura systémů pro číslicové zpracování signálů . . . . .	12
<b>3</b>	<b>Fourierovy transformace</b>	<b>14</b>
3.1	Základní diskrétní signály . . . . .	14
3.2	Fourierova transformace diskrétních signálů (FTD) . . . . .	15
3.2.1	Základní vlastnosti FTD . . . . .	18
3.2.2	Příklady spekter důležitých diskrétních signálů . . . . .	19
3.2.3	Diskrétní konvoluce . . . . .	21
3.3	Diskrétní Fourierova transformace (DFT) . . . . .	21
3.3.1	Základní vlastnosti DFT . . . . .	23
3.3.2	Volba počtu bodů N pro DFT . . . . .	26
3.3.3	Kruhová konvoluce a její užití . . . . .	29
3.4	Rychlá Fourierova transformace (FFT) . . . . .	31
3.4.1	Vyjádření DFT pomocí matic . . . . .	35
3.5	Použití DFT pro zpracování spojitych signálů . . . . .	35
3.6	Využití DFT pro spektrální analýzu periodických signálů . . . . .	36
3.6.1	Souvislost DFT s koeficienty Fourierovy řady . . . . .	39
<b>4</b>	<b>Číslicové filtry</b>	<b>41</b>
4.1	Rozdělení číslicových filtrů . . . . .	41
4.1.1	Číslicové filtry s konečnou impulsní odezvou - FIR filtry . . . . .	42
4.1.2	Číslicové filtry s nekonečnou impulsní odezvou - IIR filtry . . . . .	44
4.1.3	Rozdělení číslicových filtrů podle účelu . . . . .	49
4.1.4	Princip adaptivních číslicových filtrů a příklady jejich aplikací . . . . .	56
4.2	Metody návrhu číslicových filtrů . . . . .	57
4.2.1	Návrh filtrů s konečnou impulsní odezvou . . . . .	59
4.2.2	Návrh filtrů s nekonečnou impulsní odezvou . . . . .	66

4.3 Transpozice číslicových filtrů . . . . .	72
4.4 Porovnání vlastností FIR filtrů a IIR filtrů . . . . .	73
<b>5 Zpracování stochastických signálů</b>	<b>74</b>
5.1 Popis náhodných signálů z hlediska amplitudové struktury . . . . .	76
5.1.1 Popis amplitudového rozdělení pomocí funkčních závislostí . . . . .	76
5.1.2 Popis amplitudového rozdělení pomocí číselných hodnot . . . . .	80
5.1.3 Dvě základní spojitá rozdělení pravděpodobnosti . . . . .	84
5.2 Popis náhodných signálů v časové oblasti . . . . .	85
5.2.1 Korelační a kovarianční funkce určené ze souboru . . . . .	86
5.2.2 Korelační a kovarianční funkce určené jako střední hodnoty v čase . . . . .	87
5.2.3 Užití korelačních funkcí . . . . .	89
5.2.4 Měření korelačních funkcí . . . . .	93
5.3 Popis náhodných signálů ve frekvenční oblasti . . . . .	95
5.3.1 Způsoby získání výkonových spektrálních hustot . . . . .	99
5.3.2 Použití výkonových spektrálních hustot . . . . .	102
5.4 Měření signálů za přítomnosti šumu . . . . .	102
5.4.1 Korelační filtrace . . . . .	103
5.4.2 Průměrování signálu . . . . .	104
5.4.3 Frekvenční filtrace . . . . .	105
5.4.4 Synchronní (koherentní) detekce . . . . .	106
5.4.5 Společný popis metod zpracování signálu za přítomnosti šumu . . . . .	107
5.4.6 Vlnková transformace . . . . .	107
<b>6 Úvod části o zpracování obrazů</b>	<b>108</b>
6.1 Zpracování obrazových signálů, vztah k počítačovému vidění a počítačové grafice	108
6.2 Reprezentace obrazu a úlohy analýzy obrazu . . . . .	109
6.3 Obrazová funkce . . . . .	112
<b>7 Digitální obraz</b>	<b>114</b>
7.1 Digitalizace . . . . .	114
7.2 Vlastnosti digitálního obrazu . . . . .	115
7.3 Metrické a topologické vlastnosti obrazu . . . . .	115
7.4 Další pojmy . . . . .	118
7.5 Barevný obraz . . . . .	120
<b>8 Pořízení obrazu</b>	<b>121</b>
8.1 Obrazy jako radiometrická měření . . . . .	121
8.1.1 Radiometrické úvahy, pojmy . . . . .	121
8.1.2 Rovnice ozáření . . . . .	123
8.1.3 Odrazivost povrchu . . . . .	124
8.1.4 Rovnice ozáření a tvar ze stínování . . . . .	127
8.1.5 Tvar z fotometrického stereia . . . . .	128
8.2 Praktické otázky kolem vzniku intenzitního 2D obrazu . . . . .	130
8.2.1 Optická část . . . . .	130
8.2.2 Snímače světelné energie používané v kamerech . . . . .	133
8.2.3 Kamera . . . . .	134

<b>9 Geometrické transformace a transformace hodnot jasu</b>	<b>136</b>
9.1 Geometrické transformace . . . . .	136
9.1.1 Transformace souřadnic bodů . . . . .	137
9.1.2 Aproximace jasové funkce . . . . .	138
9.2 Transformace hodnot jasu . . . . .	141
9.2.1 Jasové korekce . . . . .	141
9.2.2 Transformace jasové stupnice . . . . .	141
<b>10 Diskrétní lineární integrální transformace</b>	<b>145</b>
10.1 Základní teorie . . . . .	145
10.2 Fourierova transformace . . . . .	146
10.3 Hadamardova transformace . . . . .	149
10.4 Diskrétní kosinová transformace . . . . .	150
10.5 Vlnková transformace (wavelets) . . . . .	150
10.6 Další lineární integrální transformace . . . . .	152
10.7 Použití lineárních diskrétních transformací pro zpracování obrazů . . . . .	152
<b>11 Filtrace šumu, poruch a detekce hran</b>	<b>157</b>
11.1 Statistický princip filtrace šumu . . . . .	157
11.2 Lokální filtrace . . . . .	157
11.3 Lokální vyhlazování obrazu . . . . .	158
11.3.1 Lineární metody vyhlazování . . . . .	159
11.3.2 Nelineární metody vyhlazování . . . . .	161
11.4 Hledání hran . . . . .	163
11.4.1 Pojem hrany a princip gradientních operátorů . . . . .	163
11.4.2 Jednoduché konvoluční masky approximující derivace obrazové funkce . . . . .	166
11.4.3 Hrany jako průchody nulou druhé derivace obrazové funkce . . . . .	168
11.4.4 Volba měřítka ve zpracování obrazů . . . . .	170
11.4.5 Cannyho hranový detektor . . . . .	171
11.5 Obnovení obrazu při známé degradaci . . . . .	173
11.5.1 Lineární model degradaci . . . . .	173
11.5.2 Matematický popis snadno odstranitelných degradací . . . . .	174
11.5.3 Obnovení obrazu inverzní filtrace nebo Wienerovou filtrace . . . . .	175
<b>12 Matematická morfologie</b>	<b>177</b>
12.1 Základní morfologické pojmy . . . . .	177
12.2 Binární matematická morfologie . . . . .	179
12.2.1 Dilatace . . . . .	179
12.2.2 Eroze . . . . .	181
12.2.3 Transformace tref či miň . . . . .	183
12.2.4 Otevření a uzavření . . . . .	183
12.3 Šedotónová matematická morfologie . . . . .	185
12.3.1 Vršek, stín, šedotónová dilatace a eroze . . . . .	185
12.3.2 Věta o homeomorfismu stínu, vlastnosti eroze, dilatace, otevření a uzavření	188
12.3.3 Transformace vrchní část klobouku . . . . .	189
12.4 Skelet a označování objektů . . . . .	190
12.4.1 Transformace zachovávající homotopický strom . . . . .	190
12.4.2 Skelet, maximální kruh . . . . .	191

## **Obsah**

12.4.3 Ztenčování, ztlušťování a homotopický skelet . . . . .	193
12.4.4 Značkování oblastí, konečná eroze . . . . .	195
12.4.5 Vzdálenostní funkce a efektivní výpočet konečné eroze . . . . .	198
12.4.6 Geodetické transformace . . . . .	199
12.4.7 Morfologická rekonstrukce . . . . .	200
12.5 Granulometrie . . . . .	203
12.6 Morfologická segmentace a rozvodí . . . . .	205
12.6.1 Segmentace částic, jejich značkování a rozvodí . . . . .	205
12.6.2 Morfologická segmentace binárních obrazů . . . . .	206
12.6.3 Šedotónová segmentace, rozvodí . . . . .	207
<b>Literatura</b>	<b>209</b>
<b>Rejstřík</b>	<b>214</b>

# 1. Úvod části o zpracování signálů (M. Sedláček)

Číslicové zpracování signálů hraje v současné době ve vědě a technice mimořádně důležitou roli. Představuje zpracování číselných posloupností číslicovou technikou. Zpracovávané číslicové posloupnosti mohou být vyrobeny v počítači, ale ve většině případů jsou získány digitalizací (vzorkováním, kvantováním a kódováním) analogového signálu. Tento signál může být jednorozměrný (nejčastěji funkce času) nebo vícerozměrný (nejčastěji dvourozměrný - funkce dvou souřadnic, která vyjadřuje dvourozměrný obraz; tímto typem signálů se zabývá druhá část tohoto skripta).

Oblast číslicového zpracování signálů se v posledních desetiletích bouřlivě rozvíjí. Jsou vyvíjeny nové algoritmy a zkoumány jejich chyby. Neustále jsou zlepšovány parametry speciálních polovodičových obvodů určených ke zpracování signálů, tzv. *číslicových signálových procesorů*. Jde o VLSI obvody obsahující aritmeticko-logickou jednotku, násobičku, generátor adres, paměti a pomocné registry, a vykonávající všechny instrukce svého jazyka (asembleru) za stejnou dobu (jeden instrukční cyklus). Instrukční cyklus současných DSP obvodů trvá např. 50ns. Poslední generace signálových procesorů pracují s plovoucí řádovou čárkou, takže programátorům odpadají problémy s přetečením mezivýsledků, a navíc je lze často programovat v některém z programovacích jazyků vyšší úrovně (např. C).

Důležitým pojmem teorie číslicového zpracování signálů je *zpracování signálů v reálném čase*. Tímto označením se zpravidla vyjadřuje situace, kdy je celý výpočet algoritmu číslicového zpracování proveden před příchodem dalšího vzorku signálu na vstup bloku blokového schématu, který vyjadřuje zpracování. Musí být tedy čas potřebný k výpočtu algoritmu zpracování kratší než vzorkovací interval. V případech, kdy se zpracovává celý blok dat najednou, jako je tomu u výpočtu rychlé Fourierovy transformace, se výpočet v reálném čase realizuje tak, že FFT spektrální analyzátor má na vstupu dvě paměti, a zatímco se algoritmem FFT zpracovává obsah jedné paměti, plní se druhá paměť. Tím je zajištěno, že se žádná data pro zpracování neztratí.

Číslicové zpracování signálů se využívá v telekomunikacích, v řídící a měřicí technice, při zpracování řeči, v seismologii, radioastronomii, v počítačové grafice, při zpracování obrazů, v lékařství, spotřební elektronice, vojenských aplikacích a dalších oborech. Přitom se využívá známých výhod zpracování signálů v číslicové formě (malá závislost na tolerancích součástek, malá citlivost na rušivé signály, volba přesnosti zpracování délhou slova, programovatelnost použitých obvodů a tím univerzálnost jejich použití).

V měřicí technice se metody zpracování signálů používají zejména při zpracování vzorkovaných dat v moderních číslicových přístrojích (analyzátorech signálů, spektrálních analyzátorech, analyzátorech energie, ale i v některých moderních číslicových multimetrech) a při zpracování vzorkovaných dat na počítačích nebo pomocí číslicových signálových procesorů. Využívají se zejména algoritmy rychlé Fourierovy transformace a číslicové filtrace, které jsou součástí řady programů pro výpočet frekvenčních spekter signálů, výkonových spektrálních

hustot, korelačních, kovariančních a cohorečních funkcí apod. Knihovny pro zpracování signálů jsou součástí řady komerčně dostupných programových balíků pro počítače. K úlohám zpracování signálů patří analýza signálů a nalezení různých parametrů signálů, úprava signálů pro získání požadovaných vlastností a generování signálů s požadovanými vlastnostmi.

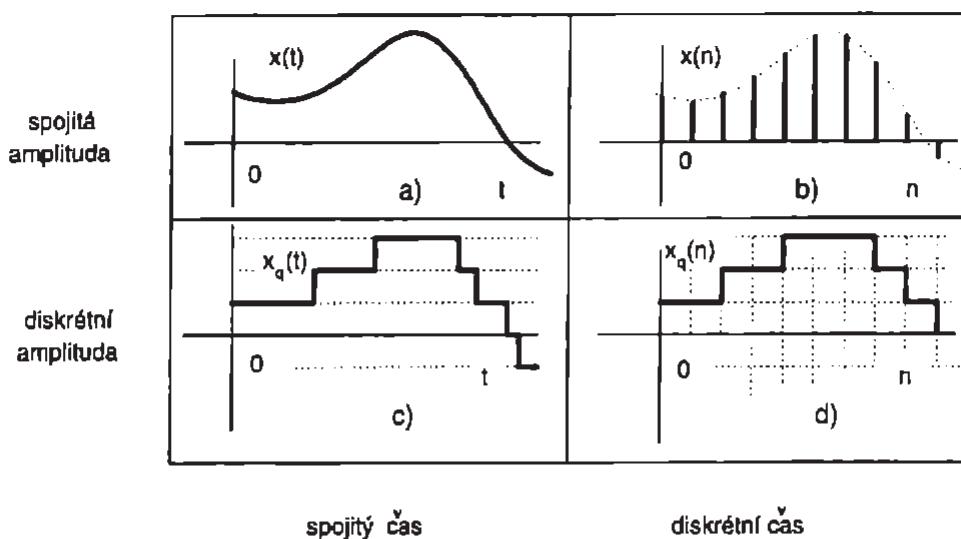
Moderní *analyzátory signálu* umožňují měřit řadu funkcí a používají složité algoritmy zpracování signálu. Pokud uživatel vlastnosti a způsob výpočtu měřených funkcí nezná, může výsledky měření interpretovat zcela nesprávně.

Účelem této první části skripta je seznámit čtenáře se základními algoritmy číslicového zpracování signálů a se základními metodami zpracování stacionárních a jednorozměrných stochastických signálů.

Na závěr této úvodní partie uvedeme *rozdělení signálů*. Z hlediska determiničnosti signály dělíme na *signály deterministické* a *signály náhodné stochastické*.

*Deterministické signály* lze popsat matematickou funkcí, takže umíme určit jejich hodnotu pro libovolnou hodnotu nezávisle proměnné (kterou bývá u jednorozměrných signálů nejčastěji čas). Patří k nim *signály periodické*, *signály kvaziperiodické* a *signály přechodné*. Podskupinou periodických signálů jsou signály *antiperiodické*, pro které platí  $f(t + T_p/2) = -f(t)$ , takže jejich druhá polovina periody je podle vodorovné osy osové souměrná s první polovinou posunutou o polovinu periody. *Kvaziperiodické signály* jsou tvořeny součtem několika harmonických (sinusových) signálů, jejichž frekvence nejsou celistvými násobky jedné základní frekvence. Jejich příkladem jsou amplitudově nebo frekvenčně modulované signály. *Přechodné signály* trvají omezenou dobu nebo teoreticky neomezenou dobou, ale periodicky se neopakují. Jejich kmitočtové spektrum je spojité. Příkladem jsou průběhy napětí při přechodných dějích v elektrických obvodech.

*Náhodné (stochastické) signály* je možno popisovat pouze statisticky. Zpracování náhodných signálů je věnován text na str. 74 tohoto skripta.



Obrázek 1.1 Rozdělení signálů z hlediska spojitosti v obou osách

Z hlediska *spojitosti v čase a v amplitudě* popisuje signály obr. 1.1. Signály v obr. 1.1 a jsou tzv. *analogové signály*, které se vyskytují v analogových obvodech a jsou spojitymi funkcemi argumentu  $t$  definovanými pro libovolné  $t$ . V praxi jde o reálné funkce, jejichž argumentem je čas.

Signály z obr. 1.1 b jsou *signály diskrétní v čase*. Mohou nabývat libovolné hodnoty amplitudy v určitém pásmu. Jejich teorie je zpracována ve většině prací týkajících se „číslicového zpracování signálů“. Signály tohoto typu zpracovávají obvody s nábojovou vazbou (CCD - charge-coupled devices) a filtry se spínanými kapacitory (SCF - switched-capacitor filters).

Signály z obr. 1.1 c jsou *víceúrovňové signály spojité v čase*. V praxi se vyskytují většinou pouze jako dvouúrovňové nebo trojúrovňové.

Signály z obr. 1.1 d jsou *vzorkované číslicové signály*. V praxi se většinou získávají digitalizací (tj. vzorkováním, kvantováním v úrovni a kódováním) analogových signálů. S tímto druhem signálů pracují všechny číslicové počítače, včetně signálových procesorů. Počet amplitudových úrovní signálu je určen počtem bitů použitého analogově-číslicového převodníku, případně délkom registrů použitého počítače. V teorii se kvantování signálu zohlední zavedením generátorů kvantovacího šumu do blokových schémat popisujících systémy pracující se signály z obr. 1.1 b, a to v místech, kde ke kvantování dochází.

Signály diskrétní v čase (z obr. 1.1 b, d) jsou z matematického hlediska posloupnosti. Přechod od analogových signálů k těmto posloupnostem se provádí operací zvanou *vzorkování*, při kterém je nutno dodržet vzorkovací větu.

## 2. Vzorkování signálu a rekonstrukce vzorkovaného signálu

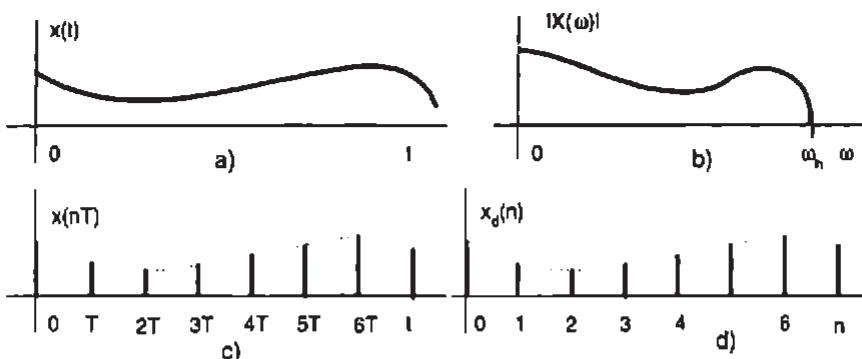
### 2.1 Vzorkovací věta

Číslicové signály používané v měření a řízení jsou v praxi většinou získávány vzorkováním analogového signálu. Získané vzorky jsou pomocí analogově-číslicového převodníku převedeny na čísla (jsou kvantovány v úrovni a vyjádřeny v číselném kódu používaném v daném analogově-číslicovém převodníku). V některých případech se převod na čísla provede až po předzpracování nekvantovaných vzorků (např. v zařízeních využívajících obvody s nábojovou vazbou a ve filtroch se spínánými kapacitory). Vzorkováním se signálu  $x(t)$  (spojitému v čase) přiřazuje signál  $x_d(n)$  diskrétní v čase (posloupnost), pro který platí

$$x_d(n) = x(nT). \quad (2.1)$$

Zde  $T$  je *vzorkovací interval* (*vzorkovací perioda*) a index  $d$  vyjadřuje diskrétnost signálu. Vzorkování je prováděno s frekvencí

$$f_{vz} = \frac{1}{T}. \quad (2.2)$$



Obrázek 2.1 a) Analogový signál  $x(t)$ , b) konečné frekvenční spektrum tohoto signálu  $|X(\omega)|$ , c) vzorkovaný signál  $x(nT)$ , d) diskrétní signál  $x_d(n)$ .

Vzorkování s pevnou vzorkovací frekvencí je v praxi nejčastější a nazývá se periodické nebo *sekvenční vzorkování*. Jinými typy jsou vzorkování náhodné, při kterém jsou vzorkovací okamžiky rozděleny náhodně podél časové osy, a vzorkování adaptivní, u kterého se frekvence vzorkování mění podle určitého kritéria (např. podle toho, zda dochází ke změně hodnoty signálu). *Vzorkovací věta*, za jejíž autory jsou uváděni nejčastěji americký matematik Claude Elwood Shannon a ruský radiotechnik Vladimír Alexandrovič Kotelníkov, stanovuje podmínu,

za které se při vzorkování signálu neztratí žádná informace. Je možné ji formulovat např. následujícím způsobem:

**Věta 2.1 Vzorkovací.** Pokud signál  $x(t)$  spojitý v čase obsahuje pouze frekvenční složky s frekvencemi menšími než  $f_{max}$ , pak  $x(t)$  může být jednoznačně rekonstruován z posloupnosti ekvidistantních vzorků  $x(nT)$ , pokud vzorkovací frekvence  $f_{vz} = 1/T$  je větší než  $2f_{max}$ , čili

$$f_{vz} > 2f_{max}. \quad (2.3)$$

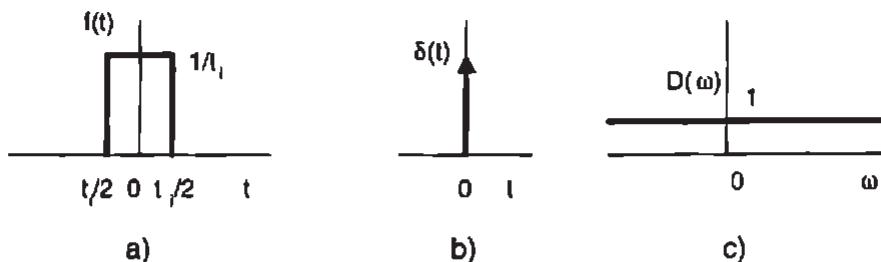
▲

Signál obsahující pouze frekvenční složky s frekvencemi menšími než konečné číslo  $f_{max}$  se nazývá *frekvenčně omezený signál*, případně signál s omezeným frekvenčním spektrem (angl. band-limited signal). Frekvence  $f_{vz}/2$  je *Nyquistova frekvence*.

Pro signál splňující větu lze tedy z hodnot vzorků  $x(nT)$  vypočítat hodnotu původního analogového signálu  $x(t)$  pro libovolné  $t$ . Postup k tomu použitý se nazývá *rekonstrukce signálu* a tři metody rekonstrukce budou popsány dále.

## 2.2 Diracův impuls a jeho vlastnosti

V teorii spojitého signálu a systémů má důležitou roli speciální signál, nazývaný *Diracův impuls*,  $\delta$  – *impuls*, případně  $\delta$  – *funkce*. Jde o signál v praxi realizovatelný pouze přibližně. Jeho délka je nekonečně malá a amplituda nekonečně velká, přičemž jeho plocha je rovná jedné (proto se někdy nazývá také *jednotkový impuls*). Z matematického hlediska jde o zobecněnou funkci neboli distribuci. Podrobnější výklad teorie distribucí přesahuje rámec těchto skript. Řadu informací najde zájemce v [72]. Zde budou uvedeny pouze základní informace, které umožní zjednodušit některá dále uvedená odvození. Zejména využijeme toho, že násobení signálu posloupností Diracových impulů je ekvivalentní vzorkování signálu v okamžicích těchto impulů.



**Obrázek 2.2** Diracův impuls: a) obdélníková vytvářející funkce, b) symbol Diracova impulsu, c) spektrum Diracova impulsu.

Diracův impuls lze definovat jako limitu posloupnosti určitých funkcí  $f_i(t)$  pro délku těchto funkcí blížící se nule. Nejjednodušší funkcí použitelnou v roli vytvářející funkce je pravoúhlý impuls podle obr. 2.2 a. Pokud jde šířka tohoto impulsu k nule a přitom se zachovává jednotková plocha impulsu, stává se impuls nekonečně krátkým a nekonečně vysokým. Platí pro něj tedy

$$\delta(t) = \lim_{t_i \rightarrow 0} f_i(t) = \begin{cases} \infty, & t = 0, \\ 0, & t \neq 0, \end{cases} \quad (2.4)$$

a současně

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1. \quad (2.5)$$

Pro  $\delta$ -impuls posunutý o hodnotu  $t_0$  platí

$$\delta(t - t_0) = \lim_{t_i \rightarrow t_0} f(t - t_0) = \begin{cases} \infty, & t = t_0, \\ 0, & t \neq 0, \end{cases} \quad (2.6)$$

a současně

$$\int_{-\infty}^{+\infty} \delta(t - t_0) dt = 1. \quad (2.7)$$

Z výše uvedených vztahů plyne *vzorkovací vlastnost  $\delta$ -impulu*:

$$\int_{-\infty}^{+\infty} \delta(t - t_0) x(t) dt = x(t_0) \int_{-\infty}^{+\infty} \delta(t - t_0) dt = x(t_0), \quad (2.8)$$

kde jsme použili vztah (2.13). Protože je  $\delta(t)$  sudá funkce, platí také

$$\int_{-\infty}^{+\infty} \delta(t_0 - t) x(t) dt = x(t_0). \quad (2.9)$$

Chceme-li vypočítat spektrum funkce  $\delta(t)$ , dosadíme do definičního vztahu pro *přímou Fourierovu transformaci*

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \quad (2.10)$$

za funkci  $x(t)$  obdélníkovou vytvářející funkci dle obr. 2.2 a. Výpočtem integrálu a limity (výpočet l'Hospitalovým pravidlem) dostaneme

$$D(\omega) = \lim_{t_i \rightarrow 0} \frac{\sin(\omega t_i / 2)}{(\omega t_i / 2)} = 1. \quad (2.11)$$

Diracův impuls budeme značit šipkou dle obr. 2.2 b. Pokud bude váha (plocha) impulsu odlišná od 1, bude její hodnota vyjádřena číslem u šipky nebo naznačena délkou šipky. Použijeme-li pro vyjádření souvislosti mezi předmětem a obrazem Fourierovy transformace symbol  $\longleftrightarrow$ , můžeme psát

$$\delta(t) \longleftrightarrow 1 \quad (2.12)$$

Odtud plyne, že spektrum Diracova impulsu je reálné a má tedy nulovou fázii.

Další důležité vlastnosti Diracova impulsu jsou

$$x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0), \quad (2.13)$$

$$x(t) * \delta(t - t_0) = x(t - t_0), \quad (2.14)$$

$$a\delta(t) \longleftrightarrow a, \quad (2.15)$$

$$\delta(t - t_0) \longleftrightarrow e^{-j\omega t_0}, \quad (2.16)$$

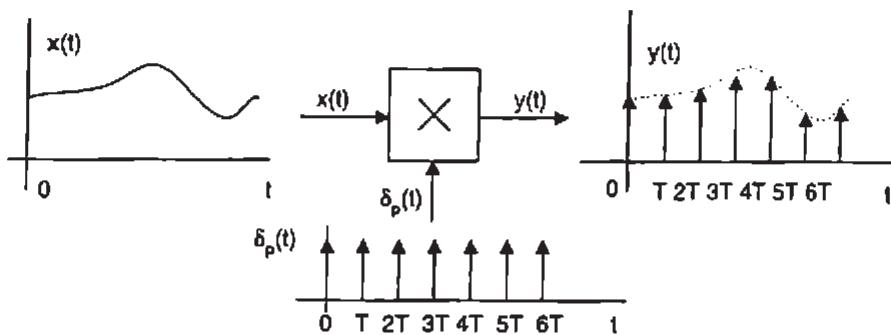
$$\delta(at) = \frac{1}{|a|}\delta(t). \quad (2.17)$$

Důležitá je zejména skutečnost, že konvoluci libovolné funkce s Diracovým impulsem posunutým do bodu  $t_0$  je tato funkce posunutá o stejnou hodnotu. Pro nulový posuv odtud vyplývá, že konvoluci určité funkce s Diracovým impulsem je nezměněná původní funkce:

$$x(t) * \delta(t) = x(t). \quad (2.18)$$

Význam Diracova impulsu spočívá mimo jiné v tom, že odezvu lineárního spojitého systému na tento impuls je tzv. *impulsní odezva (váhová funkce)  $h(t)$*  a konvolucí impulsní odezvy systému a vstupního signálu je výstupní signál systému.

## 2.3 Odvození průběhu spektra vzorkovaného signálu



Obrázek 2.3 Vzorkování signálu vyjádřené pomocí Diracových impulsů

Vzorkování signálu lze z matematického hlediska vyjádřit jako násobení signálu posloupnosti Diracových impulsů, jak to ukazuje obr. 2.3 (viz např. [86]).

Pro výstupní signál vzorkovače  $y(t)$  platí

$$\begin{aligned} y(t) &= x(t) \cdot \delta_p(t) = x(t) \cdot \sum_{i=-\infty}^{\infty} \delta(t - iT) = \\ &= \sum_{i=-\infty}^{\infty} x(t) \cdot \delta(t - iT) = \sum_{i=-\infty}^{\infty} x(iT) \cdot \delta(t - iT), \end{aligned} \quad (2.19)$$

neboť  $x(t)$  nezávisí na  $i$  a platí (2.13).

Pro nalezení spektra tohoto signálu využijeme vztahu pro spektrum posloupnosti Diracových impulsů  $\delta_p(t)$  (odvození viz např. [86])

$$\sum_{i=-\infty}^{\infty} \delta(t - iT) \longleftrightarrow \omega_{vz} \cdot \sum_{k=-\infty}^{\infty} \delta(\omega - k \cdot \omega_{vz}), \quad (2.20)$$

čili

$$\delta_p(t) \longleftrightarrow \omega_{vz} \cdot \delta_p(\omega). \quad (2.21)$$

Obrazem posloupnosti Diracových impulsů s periodou  $T = 1/f_{vz}$  je ve frekvenční oblasti opět



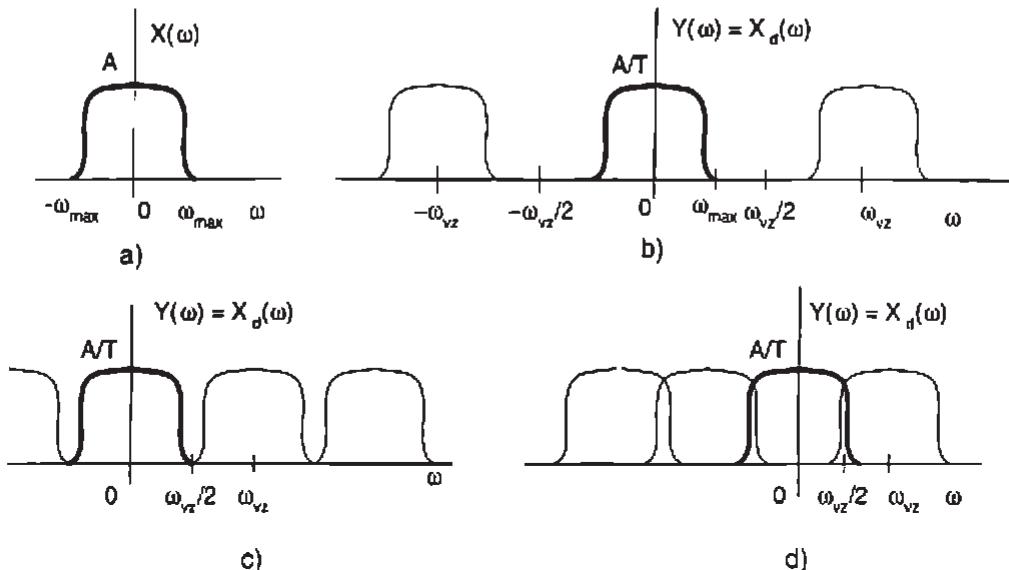
Obrázek 2.4 Posloupnost Diracových impulsů a její spektrum

posloupnost Diracových impulsů, ale s periodou  $\omega_{vz}$  a násobená konstantou  $\omega_{vz}$  (obr. 2.4). Spektrum průběhu  $y(t)$  z obr. 2.3 dostaneme tedy tímto výpočtem:

$$Y(\omega) = \frac{1}{2\pi} X(\omega) * \omega_{vz} \cdot \sum_{k=-\infty}^{\infty} \delta(\omega - k \cdot \omega_{vz}) =$$

$$\begin{aligned}
 &= \frac{1}{T} \sum_{k=-\infty}^{\infty} X(\omega) * \delta(\omega - k \cdot \omega_{vz}) = \\
 &= \frac{1}{T} \sum_{k=-\infty}^{\infty} X(\omega - k \cdot \omega_{vz}). \tag{2.22}
 \end{aligned}$$

Je tedy spektrum vzorkovaného signálu tvořeno periodickým opakováním spektra původního spojitého signálu, vynásobeného hodnotou  $1/T$ . PerIODA opakování spektra je rovna vzorkovací frekvenci (případně kruhové vzorkovací frekvenci, pokud pracujeme s kruhovými frekvencemi), viz obr. 2.5.



Obrázek 2.5 Souvislost mezi spektrem analogového signálu  $x(t)$  (a) a spektrem vzorkovaného signálu  $x_s(nT)$  (b); obr. c znázorňuje případ  $\omega_{max} = \omega_{vz}/2$ , obr. d případ nesplnění vzorkovací věty ( $\omega_{max} > \omega_{vz}/2$ )

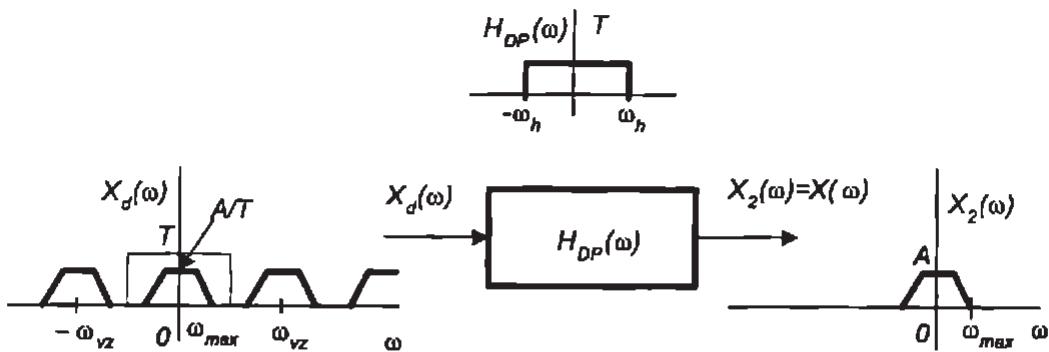
## 2.4 Rekonstrukce spojitého signálu z posloupnosti vzorků

### 2.4.1 Rekonstrukce pomocí ideální dolnofrekvenční propusti

Je-li při vzorkování signálu splněna vzorkovací věta, odpovídá průběh spekter před a po vzorkování obr. 2.5 a, b, c. Pokud tedy vzorkovaný signál necháme projít filtrem typu ideální dolnofrekvenční (dolní) propust se zesílením  $T$  a mezní frekvencí rovnou  $\omega_h = \omega_{vz}/2$ , čili filtrem s přenosem

$$H_{DP}(\omega) = \begin{cases} T, & |\omega| \leq \omega_h, \\ 0, & |\omega| > \omega_h, \end{cases} \tag{2.23}$$

dostaneme na výstupu filtru původní signál - viz obr. 2.6. Pokud ale vzorkovací věta splněna není, dochází ve frekvenční oblasti k překrývaní opakovaných spekter, spektrum ve frekvenčním pásmu  $(-\omega_{vz}/2, \omega_{vz}/2)$  je deformováno a rekonstrukcí vzorkovaného signálu není možno získat původní signál - viz obr. 2.5 d. Tomuto jevu se anglicky říká *aliasing* a v praxi je třeba mu zabránit, např. zapojením analogové dolnofrekvenční propusti s mezní frekvencí rovnou  $\omega_{vz}/2$  na vstup řetězce pro zpracování signálu.



Obrázek 2.6 Princip rekonstrukce spojitého signálu ze signálu vzorkovaného pomocí ideální dolní propusti

Interpolaci vzorec (vztah pro výpočet hodnot spojitého signálu i mezi okamžiky vzorkování) odvodíme s využitím základních vlastností Diracova impulsu. Impulsní odezva ideální dolní propusti je zpětnou Fourierovou transformací vztahu (2.23). Je tedy dáná vztahem

$$h_{DP}(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \quad (2.24)$$

(Při zpětné Fourierově transformaci stačí integrovat v mezích  $-\omega_{vz}/2, \omega_{vz}/2$ . Přitom platí  $\omega_{vz}/2 = \pi/T$ .) Z (2.24) plyne  $h_{DP}(0) = 1$  a  $h_{DP}(nT) = 0$  pro  $n$  celé a  $n \neq 0$ . Pro výstupní signál dolní propusti z obr. 2.6 platí

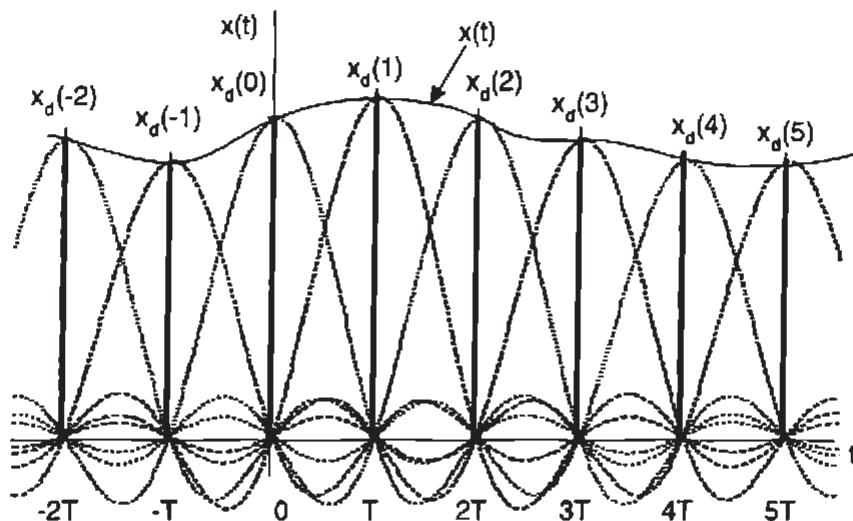
$$\begin{aligned} x_2(t) &= h_{DP}(t) * x_d(n) = \frac{\sin(\pi t/T)}{\pi t/T} * \sum_{n=-\infty}^{\infty} x(nT) \cdot \delta(t - nT) = \\ &= \sum_{n=-\infty}^{\infty} x(nT) \cdot \left\{ \frac{\sin(\pi t/T)}{\pi t/T} * \delta(t - nT) \right\}. \end{aligned} \quad (2.25)$$

Využitím vztahu (2.14) dostaneme konečnou podobu interpolaciho vzorce

$$x_2(t) = x(t) = \sum_{n=-\infty}^{\infty} x(nT) \cdot \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T}. \quad (2.26)$$

Způsob interpolace pro nalezení hodnot  $x(t)$  pro  $t \neq nT$  je zřejmý z obr. 2.7, kde je vidět, že pro  $t = nT$  jsou hodnoty rekonstruované funkce totožné s hodnotami  $x(nT)$ , protože příspěvky všech posunutých funkcí  $(\sin x)/x$  jsou v těchto bodech nulové.

Popsaná rekonstrukce pomocí ideální dolní propusti se nazývá *Whittakerova rekonstrukce* nebo *Shannonova interpolace*. Ve skutečnosti ji ale nelze podle vztahu (2.26) realizovat, protože suma má nekonečný počet členů. V praxi se proto realizuje přibližně, s využitím konečného počtu členů řady (např. deseti). Ani po tomto zjednodušení nelze tuto rekonstrukci využít pro zpracování v reálném čase, protože ideální dolní propust s frekvenční charakteristikou dle obr. 2.6 není kauzální (její impulsní odezva je nenulová pro záporné časy). Pokud se ale pracuje s konečným počtem členů řady, je možno impulsní odezvu posunout doprava tak, aby se filtr stal kauzálním. Důsledkem tohoto posuvu je nenulovost fázové frekvenční charakteristiky, což plyne z věty o posuvu v předmětu ve Fourierově transformaci a toho, že Fourierovým obrazem reálné a sudé funkce je opět reálná a sudá funkce (viz kap. 3).



Obrázek 2.7 Rekonstrukce spojitého signálu ze signálu vzorkovaného pomocí ideální dolní propusti

#### 2.4.2 Rekonstrukce pomocí vzorkovače s pamětí (Lagrangeova interpolátoru řádu 0)

V praxi je nejčastěji vzorkovaný průběh (posloupnost čísel) převáděn na spojitý signál pomocí číslicově-analogového převodníku, který na svém výstupu drží prakticky konstantní hodnotu napětí až do příchodu dalšího vzorku (do spuštění dalšího převodu). Použitý obvod se nazývá Lagrangeův interpolátor řádu nula, anglicky také „sample-and-hold“, „zero-order hold“, nebo „boxcar“. Jeho impulsní odezva je popsána vztahem

$$h_0(t) = \begin{cases} 1, & 0 \leq t \leq T, \\ 0, & jinde, \end{cases} \quad (2.27)$$

čili jde o obdélník o výšce 1 začínající v  $t = 0$  a končící v  $t = T$  ( $T$  je délka vzorkovacího intervalu). Výstupní funkce tohoto obvodu je schodovitá funkce, pro kterou platí

$$\hat{x}(t) = x_d(n) * h_0(t). \quad (2.28)$$

Fourierovým obrazem (frekvenčním spektrem) obdélníkového impulsu o amplitudě  $A$  a délce  $t_i$ , začínajícího v  $t = 0$ , je (viz např.[86])

$$X(\omega) = e^{-j\omega t_i/2} \cdot At_i \frac{\sin(\omega t_i/2)}{\omega t_i/2}. \quad (2.29)$$

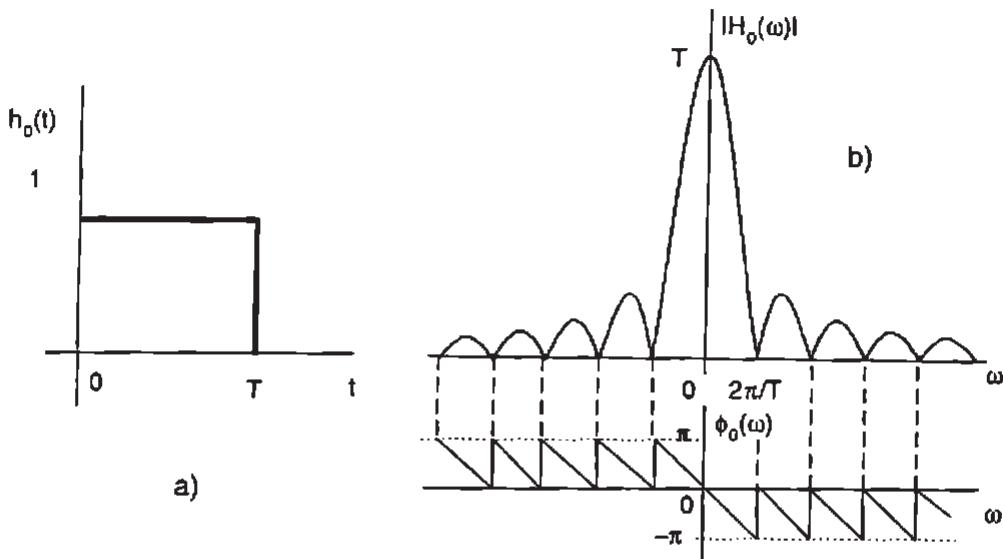
Proto lze spektrum (frekvenční charakteristiku) vzorkovače s pamětí vyjádřit vztahem

$$H_0(\omega) = e^{-j\omega T/2} \cdot T \cdot \frac{\sin(\omega T/2)}{\omega T/2}. \quad (2.30)$$

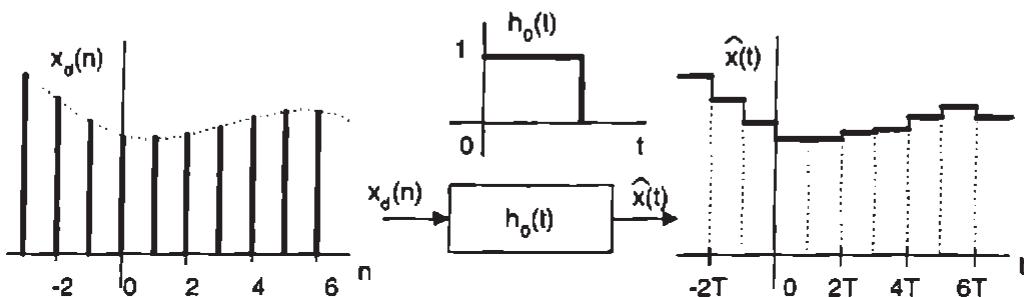
Průběh impulsní odezvy a amplitudového i fázového spektra tohoto interpolátoru uvádí obr. 2.8. Rekonstrukční vztah realizovaný tímto interpolátorem je

$$\hat{x}(t) = x_d(n-1), \quad t \in ((n-1)T, nT]. \quad (2.31)$$

Tento interpolátor lze užít pro rekonstrukci v reálném čase. Schodovitou interpolaci vzorkovaného signálu znázorňuje obr. 2.9.



Obrázek 2.8 Impulsní odezva a frekvenční spektrum vzorkovače s pamětí řádu nula

Obrázek 2.9 Rekonstrukce signálu pomocí schodovité funkce a průběh impulsní odezvy  $h_0(t)$  interpolátoru

### 2.4.3 Rekonstrukce pomocí lineárního interpolátoru

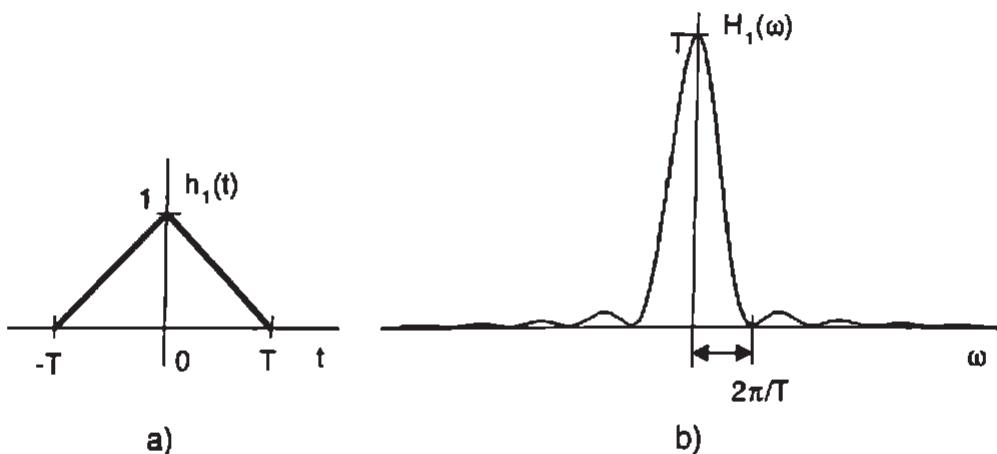
Impulsní odezva lineárního interpolátoru  $h_1(t)$  je znázorněna v obr. 2.10 a. Frekvenční charakteristika lineárního interpolátoru se získá Fourierovou transformací impulsní odezvy. Tu lze provést výhodně po zderivování impulsní odezvy, dvojím využití vztahu (2.29) a věty o obrazu integrálu ve Fourierově transformaci. Spektrum je reálná funkce  $\omega$  daná výrazem

$$H_1(\omega) = T \cdot \left[ \frac{\sin(\omega T/2)}{\omega T/2} \right]^2. \quad (2.32)$$

Jeho průběh uvádí obr. 2.10b. Rekonstruovaný signál je tvořen úsečkami spojujícími sousední hodnoty vzorků  $x(nT)$ , čili jde o lineární interpolaci. Rekonstrukční předpis tohoto interpolátoru je

$$\hat{x}(t) = x_d(n-1) \frac{nT - t}{T} + x_d(n) \frac{t - (n-1)T}{T}, \quad t \in ((n-1)T, nT). \quad (2.33)$$

Pokud chceme lineární interpolátor využít pro zpracování signálu v reálném čase, je nutno posunout impulsní odezvu doprava tak, aby byla nulová pro záporná  $t$ .



Obrázek 2.10 Impulsní odpověď a frekvenční odpověď lineárního interpolátoru

#### 2.4.4 Chyba rekonstrukce signálu

Byla-li při vzorkování signálu splněna vzorkovací věta, je rekonstrukce pomocí ideální dolní propusti bez chyby. Při použití Lagrangeových interpolátorů řádu nula nebo jedna ale dochází při rekonstrukci signálu k dynamickým chybám, které můžeme vyjádřit v časové oblasti vztahem

$$\Delta(t) = \hat{x}(t) - x(t), \quad (2.34)$$

kde  $\hat{x}(t)$  je průběh rekonstruované funkce a  $x(t)$  je původní analogový průběh před vzorkováním (a tedy i průběh ideálně rekonstruovaného signálu).

Ve frekvenční oblasti se chyba rekonstrukce projeví deformačí spektra rekonstruovaného signálu proti spektru původního signálu. Spektrum rekonstruovaného signálu dostaneme jako součin spektra vzorkovaného signálu  $X_d(e^{j\omega})$  a frekvenční odpovědi použitého rekonstrukčního filtru  $H_{RF}(\omega)$ , tedy ze vztahu

$$\hat{X}(\omega) = X_d(e^{j\omega}) \cdot H_{RF}(\omega). \quad (2.35)$$

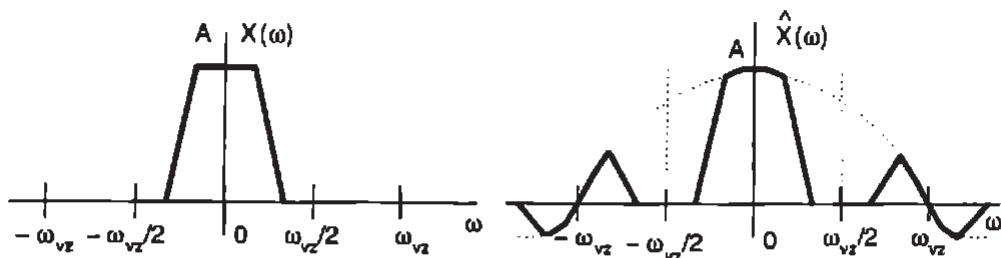
Fourierovou transformací vztahu (2.34) a užitím rovnice (2.35) získáme průběh chyby rekonstrukce (interpolace) ve frekvenční oblasti

$$\Delta(\omega) = \hat{X}(\omega) - X(\omega) = X_d(e^{j\omega}) \cdot H_{RF}(\omega) - X(\omega). \quad (2.36)$$

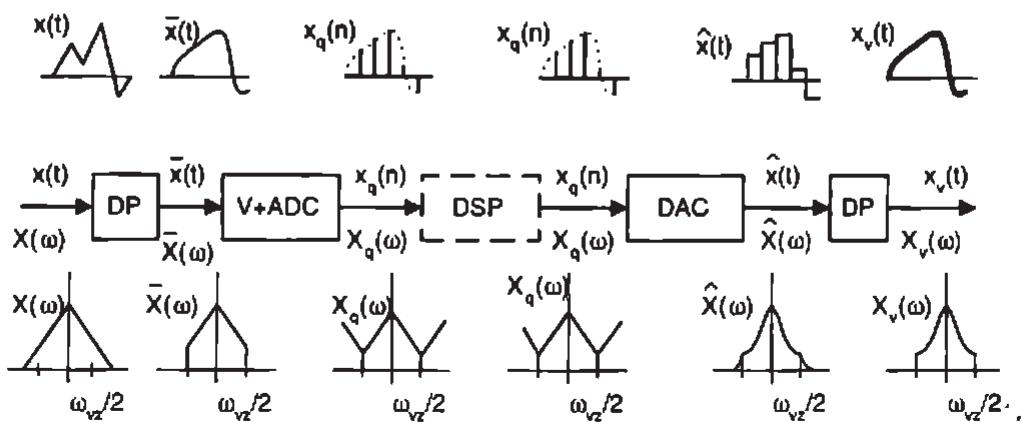
Ze vztahů (2.30) a (2.32) je zřejmé, že se pro tyto filtry liší ve frekvenčním pásmu  $(-\omega_{vz}/2, \omega_{vz}/2)$  rekonstruované spektrum  $\hat{X}(\omega)$  od spektra vzorkovaného signálu  $X_d(e^{j\omega})$ . Graficky je tato skutečnost znázorněna pro rekonstrukční filtr řádu nula v obr. 2.11. Zmenšení této chyby lze dosáhnout filtrací rekonstruovaného signálu filtrem typu dolní propust se zesílením 1 a mezni frekvencí  $\omega_{vz}/2$ , který odstraní části spektra mimo pásmo  $(-\omega_{vz}/2, \omega_{vz}/2)$ . V časové oblasti se tato filtrace projeví zmenšením sklonu skokových změn v původním rekonstruovaném signálu (vyhlazením schodovitého signálu).

## 2.5 Bloková struktura systémů pro číslicové zpracování signálů

Ve většině případů v praxi je signál určený ke zpracování signál analogový. Proto je nutno ho nejprve digitalizovat, čili převést na signál diskrétní v čase (vzorkovat) a v amplitudě (kvantovat a kódovat). Tyto operace provede analogovo-číslicový převodník a jeho řídicí obvod určující



Obrázek 2.11 Deformace spektra  $\hat{X}(\omega)$  interpolovaného signálu  $\hat{x}(t)$  při schodové interpolaci vzorkovaného signálu  $x_d(n)$ .  $X(\omega)$  je spektrum analogového signálu před vzorkováním



Obrázek 2.12 Blokové schéma systému pro číslicové zpracování signálů a příklady průběhů v časové a frekvenční oblasti

okamžiky zahájení převodů. Pro vyloučení aliasingu (chyby překládání frekvenčních pásem, způsobené nedodržením vzorkovací věty) bývá na vstupu systému zařazen analogový filtr typu dolní propust (DP, dnes často filtr se spínanými kapacitami) a mezní frekvenci  $\omega_{vz}/2$  nastavovanou někdy automaticky podle zvolené vzorkovací frekvence. Pokud se tento *antialiasingový filtr* nepoužije, potom se mohou do základního pásma dostat nejen složky tvořící součást zpracovávaného signálu, ale také šumové složky s velmi vysokými frekvencemi, které se k signálu superponují nekontrolovatelným způsobem.

Ideální dolní propust je v praxi nerealizovatelná, v praxi je signál použitym filtrem vždy pozměněn. Mezní frekvence antialiasingového filtru se volí dvakrát až třikrát nižší než  $\omega_{vz}/2$ , podle typu a řádu filtru. Obecně platí, že signál za antialiasingovým filtrem pro  $\omega > \omega_{vz}/2$  by měl být pod rozlišovací schopností A/D převodníku (ADC).

Po zpracování je v řadě případů diskrétní signál opět převáděn na analogový. K tomuto převodu slouží číslicově-analogový převodník (DAC) a dolní propust na jeho výstupu. Celý systém včetně příkladů průběhů v jeho jednotlivých místech znázorňuje obr. 2.12. Pro průběhy v obr. 2.12 se předpokládá, že blok číslicového zpracování signálů (DSP) ponechává signál bez ziměny, aby bylo možné porovnat rozdíl mezi vstupním a výstupním signálem řetězce způsobený vlivem ostatních bloků.

V teorii číslicových signálů se blok ADC respektuje zavedením zdrojů *kvantovacího šumu* do příslušného místa blokového schématu algoritmu zpracování. Obdobně se zkoumá i vliv konečné délky slova (délky použitých registrů) počítače nebo signálového procesoru. Důsledkem je nelinearity systému, která může vést k řadě neočekávaných projevů (viz např. [86, 39, 68]).

### 3. Fourierovy transformace

#### 3.1 Základní diskrétní signály

Diskrétní signály budeme pro stručnost označovat  $x(n)$ , tj. vypustíme index  $d$  ze vztahu (2.1). Pokud budeme chtít zdůraznit hodnotu vzorkovacího intervalu  $T$ , použijeme označení  $x(nT)$ . Základními diskrétními signály jsou jednotkový impuls, jednotkový skok, diskrétní harmonický signál a komplexní exponenciála. *Diskrétní jednotkový impuls*  $\delta(n)$  je definován vztahem

$$\delta(n) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0. \end{cases} \quad (3.1)$$

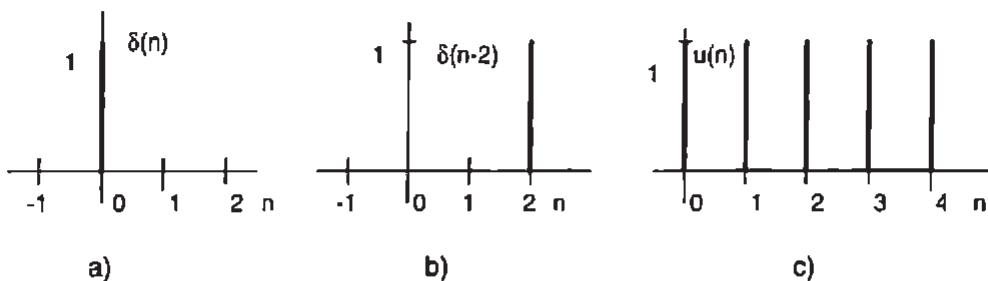
Tento signál má v teorii diskrétních signálů význam obdobný významu Diracova impulsu v teorii signálů spojitéch, ale nejde o distribuci, nýbrž o běžnou posloupnost. Pomocí posunutých jednotkových impulsů je možno vyjádřit libovolný diskrétní signál  $x(n)$  lineární kombinací:

$$x(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot \delta(n - k). \quad (3.2)$$

Jde o diskrétní obdobu Duhamelova integrálu.

Druhým základním diskrétním signálem je *diskrétní jednotkový skok*  $u(n)$ , definovaný vztahem

$$u(n) = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0. \end{cases} \quad (3.3)$$



Obrázek 3.1 Diskrétní jednotkový impuls v počátku (a) a posunutý (b); diskrétní jednotkový skok (c)

Podle (3.2) platí

$$u(n) = \sum_{k=-\infty}^{\infty} u(k) \cdot \delta(n - k) = \sum_{k=0}^{\infty} \delta(n - k). \quad (3.4)$$

I pro diskrétní signály má značný význam harmonický signál. *Diskrétní harmonický signál* je definován vztahem

$$x(n) = A \sin(\omega nT + \phi) = A \sin(\theta n + \phi), \quad (3.5)$$

kde  $\theta$  je normovaná kruhová frekvence

$$\theta = \omega T = 2\pi f / f_{vz} . \quad (3.6)$$

Signál definovaný vztahem (3.5) je periodický pouze tehdy, je-li při vzorkování analogového harmonického signálu odebrán počet vzorků  $n_p$  (celistvý) na jednu periodu nebo  $l$  vzorků na  $m$  period analogového signálu. Pak platí

$$T_{sig} = n_p \cdot T, \quad n_p - \text{přirozené}, \quad (3.7)$$

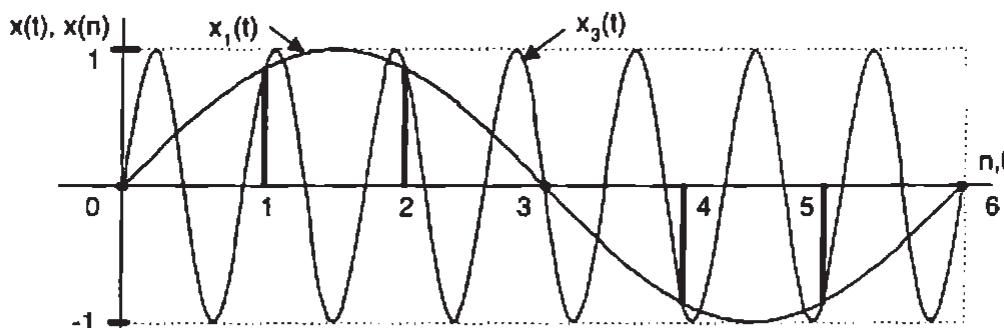
$$m \cdot T_{sig} = l \cdot T, \quad m, l - \text{přirozené}. \quad (3.8)$$

Vyjádříme-li podmínky (3.7) a (3.8) pomocí frekvence  $\theta$ , musí platit

$$2\pi/\theta = n_p, \quad n_p - \text{přirozené}, \quad (3.9)$$

$$2\pi/\theta = l/m, \quad m, l - \text{přirozené}. \quad (3.10)$$

Odtud plyne, že signál  $x_1 = \sin(\pi n/3)$  je periodický s periodou  $n_p = 6$ , signál  $x_2 = \sin(n)$  je neperiodický a signál  $x_3 = \sin((7/3)\pi n) = \sin[(\pi/3 + 2\pi)n]$  je periodický (na 7 period je 6 vzorků). Přitom signály  $x_1(n)$  a  $x_3(n)$  nelze rozlišit, protože se hodnoty  $\theta$  v jejich argumentech liší o násobek  $2\pi$ . To při vyjádření v kruhové frekvenci  $\omega$  odpovídá rozdílu o násobek  $\omega_{vz}$ . Z obr. 3.2 je zřejmé, že signál  $x_3(n)$  nesplňuje vzorkovací větu. Obr. 3.2 tak zároveň ukazuje, jak se v případě harmonického signálu projeví aliasing v časové oblasti.



Obrázek 3.2 Projev aliasingu v časové oblasti pro harmonický signál - nerozlišitelnost průběhu  $x_1(n)$  a  $x_3(n)$

Další důležitou diskrétní funkcí je komplexní exponenciála  $\exp(a + jn\theta)$ . Zvláště důležitý je případ  $a = 0$ , pro který platí Eulerův vztah

$$e^{jn\theta} = \cos(n\theta) + j \sin(n\theta) . \quad (3.11)$$

Tuto funkci je možno využít při hledání frekvenčního přenosu diskrétních systémů. Periodická je pouze tehdy, platí-li (3.9) resp. (3.10).

*Posloupnost* (diskrétní signál)  $y(n)$  zpožděná o  $n_0$  vzorků proti posloupnosti  $x(n)$  je definována vztahem

$$y(n) = x(n - n_0) . \quad (3.12)$$

## 3.2 Fourierova transformace diskrétních signálů (FTD)

Fourierova transformace je nejpoužívanější z transformací, vyjadřujících obraz signálu pomocí ortogonálních bázových funkcí (dalšími jsou např. kosinová transformace a transformace Walshova a Hadamardova). K popisu diskrétních signálů ve frekvenční oblasti se používají její

modifikace Fourierova transformace diskrétních signálů (*Fourier transform of discrete signals, FTD*), označovaná též zkratkou DTFT (z angl. Discrete-time Fourier transform) a *diskrétní Fourierova transformace, DFT*, viz str. 21. Vztah pro přímou transformaci se získá ze vztahu pro přímou Fourierovu transformaci náhradou integrálu sumou. Protože je Fourierův obraz (*frekvenční spektrum*) vzorkovaného signálu periodická funkce v  $\omega$  s periodou  $\omega_{vz} = 2\pi/T$ , zapisuje se argument obrazu ve tvaru  $\exp(j\omega T)$ . Signály v časové oblasti budeme označovat malými písmeny, jejich obrazy ve frekvenční oblasti velkými písmeny. Přímá FTD je dáná vztahem:

$$X(e^{j\omega T}) = \sum_{k=-\infty}^{\infty} x(nT) \cdot e^{-jkn\omega T} \quad (\text{FTD}). \quad (3.13)$$

Spektrum signálu je obecně komplexní funkce reálné proměnné  $\omega$  a jeho hodnoty lze tedy vyjádřit v kartézském nebo exponenciálním tvaru:

$$X(e^{j\omega T}) = R(e^{j\omega T}) + jI(e^{j\omega T}), \quad (3.14)$$

$$X(e^{j\omega T}) = A(e^{j\omega T}) \cdot e^{j\phi(e^{j\omega T})}, \quad (3.15)$$

přičemž platí

$$A = \sqrt{R^2 + I^2}; \quad (3.16)$$

$$\phi = \arg X = \arctg(\frac{I}{R}). \quad (3.17)$$

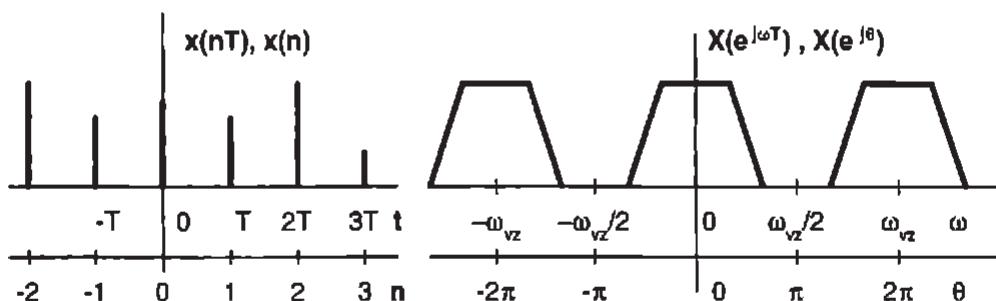
V (3.14) až (3.17) je  $R = \operatorname{Re}(X)$  a  $I = \operatorname{Im}(X)$ . Zpětná Fourierova transformace diskrétních signálů (IFTD) je definována vztahem

$$x(nT) = \frac{1}{2\pi} \int_{-\omega_{vz}/2}^{\omega_{vz}/2} T \cdot X(e^{j\omega T}) \cdot e^{jn\omega T} d\omega, \quad (3.18)$$

resp. vztahem

$$x(nT) = \frac{1}{\omega_{vz}} \int_{\omega_{vz}} X(e^{j\omega T}) \cdot e^{jn\omega T} d\omega \quad (\text{IFTD}). \quad (3.19)$$

Jde tedy o koeficienty Fourierovy řady periodické funkce  $X(e^{j\omega T})$ , periodické v  $\omega$  s periodou  $\omega_{vz}$ . Tyto koeficienty jsou v praxi nejčastěji reálné, protože pracujeme s reálnými signály v časové oblasti. Souvislost mezi signálem v časové oblasti a jeho FTD obrazem ukazuje obr. 3.3. Za základní interval ve frekvenční oblasti se obvykle považuje interval  $(-\omega_{vz}/2, \omega_{vz}/2)$ , resp v jiném vyjádření  $(-\pi/T, \pi/T)$ .



Obrázek 3.3 Souvislost diskrétního signálu a jeho spektra

V teorii diskrétních signálů se často místo s kruhovou frekvencí  $\omega$  pracuje s normovanou kruhovou frekvencí  $\theta$  (viz. (3.6)). Pomocí  $\theta$  vyjádřené spektrum má periodu  $\theta = 2\pi$  a platí pro ně

$$X(e^{j\theta}) = \sum_{k=-\infty}^{\infty} x(n) \cdot e^{-jn\theta}. \quad (3.20)$$

Pro IFTD platí

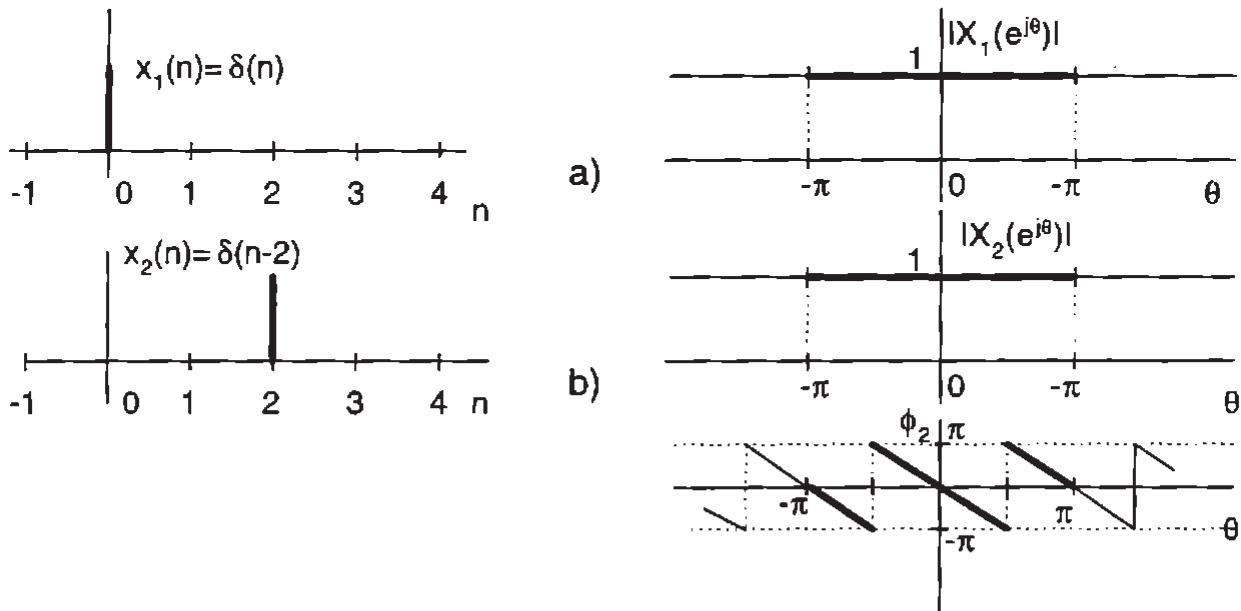
$$x(nT) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\theta}) \cdot e^{jn\theta} d\theta, \quad (3.21)$$

takže základní interval je v tomto případě  $(-\pi, \pi)$ . Tomu odpovídá spodní vodorovná osa v obr. 3.3.

Souvislost signálu a spektra budeme vyjadřovat symbolem  $\leftrightarrow$ . Budeme používat následující transformační páry:

$$x(nT) \leftrightarrow X(e^{j\omega T}), \quad (3.22)$$

$$x(n) \leftrightarrow X(e^{j\theta}). \quad (3.23)$$



Obrázek 3.4 a) Jednotkový impuls a jeho spektrum, b) posunutý jednotkový impuls a jeho spektrum

Jako příklady uvedeme spektrum jednotkového impulsu  $\delta(n)$  a posunutého jednotkového impulsu  $\delta(n-2)$ . Spektrum signálu  $\delta(n)$  je rovno 1 a má tedy nulovou fázi. Pro  $x_2(n) = \delta(n-2)$  platí

$$X_2(e^{j\theta}) = e^{-j2\theta}, \quad (3.24)$$

takže  $|X_2(e^{j\theta})| = 1$  a  $\arg(X_2(e^{j\theta})) = -2\theta$ . Tyto vztahy vyplývají přímo z 3.20. Spektra jsou zobrazena v obr. 3.4<sup>1</sup>.

<sup>1</sup>Fázová spektra se obvykle vynášejí v intervalu  $(-\pi, \pi)$ , takže se přímka jeví jako pilový průběh.

### 3.2.1 Základní vlastnosti FTD

Předpokládáme přiřazení

$$x_i(n) \longleftrightarrow X_i(e^{j\theta}). \quad (3.25)$$

1. *Linearita:*

$$k_1 x_1(n) + k_2 x_2(n) \longleftrightarrow k_1 X_1(e^{j\theta}) + k_2 X_2(e^{j\theta}). \quad (3.26)$$

V (3.26) jsou  $k_1$  a  $k_2$  libovolné konstanty. Obecněji lze říci, že obrazem lineární kombinace signálů je táž lineární kombinace odpovídajících spekter.

2. *Časový posuv* (posuv v předmětu):

$$x(n - n_0) \longleftrightarrow e^{-jn_0\theta} \cdot X(e^{j\theta}). \quad (3.27)$$

3. *Frekvenční posuv* (posuv v obrazu):

$$e^{jn\theta_0} \cdot x(n) \longleftrightarrow X(e^{j(\theta-\theta_0)}). \quad (3.28)$$

(Důkaz vztahů (3.27) a (3.28) uvádí např. [86].)

Využitím vztahů (3.28) a (3.26) lze získat obraz amplitudově modulovaného kosinusového signálu

$$2x(n) \cdot \cos(n\theta_0) = 2x(n) \cdot \frac{1}{2}(e^{jn\theta_0} + e^{-jn\theta_0}) \longleftrightarrow X(e^{j(\theta-\theta_0)}) + X(e^{j(\theta+\theta_0)}). \quad (3.29)$$

4. *Konvoluce v časové oblasti a ve frekvenční oblasti:*

$$x_1(n) * x_2(n) \longleftrightarrow X_1(e^{j\theta}) \cdot X_2(e^{j\theta}), \quad (3.30)$$

$$x_1(n) \cdot x_2(n) \longleftrightarrow \frac{1}{2\pi} X_1(e^{j\theta}) * X_2(e^{j\theta}). \quad (3.31)$$

Konvoluce v časové a frekvenční oblasti jsou přitom definovány vztahy

$$x_1(n) * x_2(n) = \sum_{i=-\infty}^{\infty} x_1(i) \cdot x_2(n-i), \quad (3.32)$$

$$X_1(e^{j\theta}) * X_2(e^{j\theta}) = \int_{-\pi}^{\pi} X_1(e^{j\lambda}) \cdot X_2(e^{j(\theta-\lambda)}) d\lambda. \quad (3.33)$$

5. *Vlastnosti spektra, je-li  $x(t)$  reálná funkce:*

Pro tento velmi důležitý praktický případ platí:

$$|X(e^{j\theta})| = |X(e^{-j\theta})|, \quad (3.34)$$

$$\arg X(e^{j\theta}) = -\arg X(e^{-j\theta}), \quad (3.35)$$

$$\operatorname{Re} X(e^{j\theta}) = \operatorname{Re} X(e^{-j\theta}), \quad (3.36)$$

$$\operatorname{Im} X(e^{j\theta}) = -\operatorname{Im} X(e^{-j\theta}), \quad (3.37)$$

$$X(e^{j\theta}) = X^*(e^{-j\theta}). \quad (3.38)$$

Je-li tedy signál tvořen reálnými vzorky, jsou amplitudové spektrum a reálná část spektra sudé funkce  $\theta$  a fázové spektrum a imaginární část spektra liché funkce  $\theta$ . Samo spektrum (komplexní funkce reálné proměnné  $\theta$ ) je tzv. *konjugovaně sudá funkce* - viz (3.38).

Protože je spektrum  $X(e^{j\theta})$  také periodické s periodou  $2\pi$ , je toto spektrum v případě reálného signálu určeno průběhem v polovině základního intervalu.

6. Je-li  $x(t)$  reálná a sudá funkce, je  $X(e^{j\theta})$  také sudá a reálná. Příklady takových funkcí jsou autokorelační funkce reálného signálu a funkce  $\cos(n\theta_0)$  (obr. 3.5).
7. Je-li  $x(t)$  reálná a lichá funkce, je  $X(e^{j\theta})$  (ryze) imaginární a lichá. Příkladem takové funkce je funkce  $\sin(n\theta_0)$  (obr. 3.5).

Těchto vlastností reálné posloupnosti vzorků  $x(n)$  lze využít, rozložíme-li tuto posloupnost na součet sudé složky  $x_s(n)$  a liché složky  $x_l(n)$ . To lze vždy provést pomocí těchto vztahů:

$$x(n) = x_s(n) + x_l(n), \quad (3.39)$$

$$x_s(n) = \frac{1}{2}(x(n) + x(-n)), \quad (3.40)$$

$$x_l(n) = \frac{1}{2}(x(n) - x(-n)). \quad (3.41)$$

### 3.2.2 Příklady spekter důležitých diskrétních signálů

Spektrum signálu  $\delta(n-2)$  uvádí rovnice (3.24). Důležité je spektrum diskrétní komplexní exponenciály, pomocí kterého lze snadno odvodit spektra diskrétního sinusového a kosinusového signálu a spektrum konstanty. Platí

$$e^{jn\theta_0} \longleftrightarrow 2\pi\delta(\theta - \theta_0). \quad (3.42)$$

Důkaz plyne přímo ze zpětné transformace tohoto spektra:

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\pi\delta(\theta - \theta_0)e^{jn\theta} d\theta = e^{jn\theta_0}. \quad (3.43)$$

Přitom jsme využili (2.8).

Odtud plynou další důležité vztahy:

$$\cos(n\theta_0) \longleftrightarrow \pi\delta(\theta - \theta_0) + \pi\delta(\theta + \theta_0), \quad (3.44)$$

$$\sin(n\theta_0) \longleftrightarrow -j\pi\delta(\theta - \theta_0) + j\pi\delta(\theta + \theta_0), \quad (3.45)$$

$$1 \longleftrightarrow 2\pi\delta(\theta). \quad (3.46)$$

Pro jejich důkaz se využije definice sinu a cosinu pomocí komplexních exponenciál a vztahu (3.42) pro  $\theta_0 = 0$ . Uvedená spektra znázorňuje obr. 3.5.<sup>2</sup>

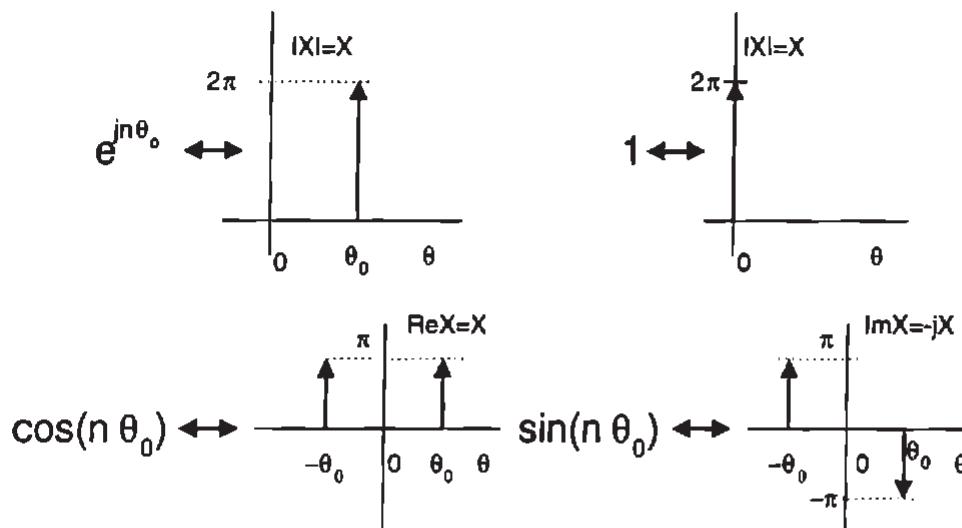
Velmi důležitým signálem je také *diskrétní obdélníkový impuls*. Jde o diskrétní obdobu signálu z obr. 2.8a a jeho průběh uvádí obr. 3.6.

Pro spektrum diskrétního obdélníka délky  $N$  (tvořeného posloupnosti  $N$  jednotkových diskrétních impulsů) platí:

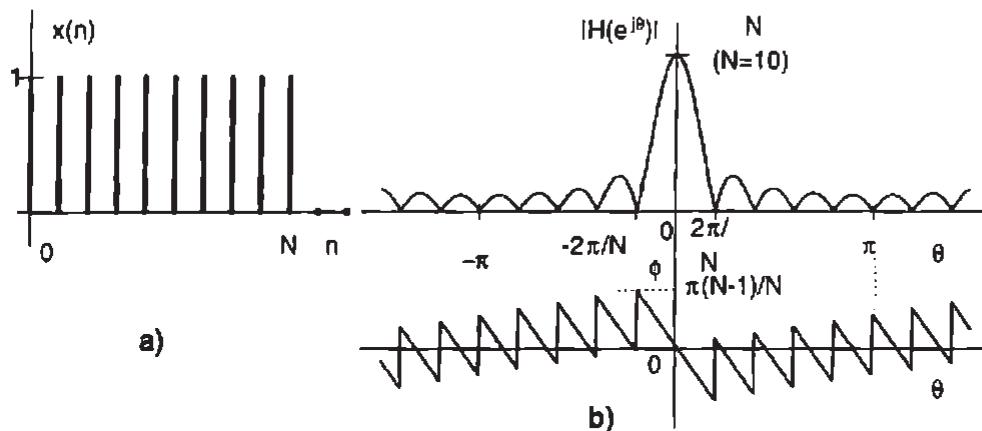
$$H(e^{j\theta}) = \sum_{n=0}^{N-1} e^{-jn\theta} = \frac{1 - e^{-j\theta N}}{1 - e^{-j\theta}} = e^{-j\theta \frac{N-1}{2}} \cdot \frac{\sin(\theta N/2)}{\sin(\theta/2)}. \quad (3.47)$$

(Ve vztahu (3.47) jsme využili vztahu pro součet  $N$  členů geometrické posloupnosti a vyjádření sinu pomocí komplexních exponenciál.) Diskrétní obdélníkový impuls podle (3.7) má spektrum

<sup>2</sup>Fázové spektrum sinusovky je nulové s výjimkou frekvencí  $-\theta_0$ , kde je rovno  $+j\pi/2$ , a  $\theta_0$ , kde je rovno  $-j\pi/2$ . To je zřejmé z (3.45).



Obrázek 3.5 FTD spektra diskrétních signálů obsahující Diracovy impulsy

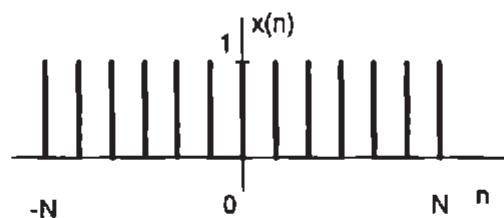


Obrázek 3.6 Kauzální diskrétní obdélníkový impuls (a) a jeho spektrum (b)

dané vztahem

$$H(e^{j\theta}) = \frac{\sin(\theta(2N+1)/2)}{\sin(\theta/2)}. \quad (3.48)$$

Tento výraz dostaneme respektováním změny délky pulsu a jeho posunu o  $-N$ . Amplitudové spektrum odpovídá obr. 3.6 po změně  $N$  za  $2N+1$ ; spektrum je reálné, takže fázové spektrum je nulové.



Obrázek 3.7 Diskrétní obdélníkový impuls

Na závěr odvodíme ještě výraz pro spektrum *kauzální mocninné funkce*

$$x(n) = a^n u(n). \quad (3.49)$$

Její spektrum pro  $|a| < 1$  je

$$H(e^{j\theta}) = \sum_{n=0}^{\infty} a^n e^{-jn\theta} = \sum_{n=0}^{\infty} (ae^{-j\theta})^n = \frac{1}{1 - ae^{-j\theta}}. \quad (3.50)$$

(Jde o součet konvergující geometrické řady.)

### 3.2.3 Diskrétní konvoluce

Obdobně jako u lineárních spojitých systémů lze i u lineárních diskrétních systémů najít výstupní signál  $y$  systému konvoluci vstupního signálu  $x$  a impulsní odezvy systému  $h$ . U diskrétních systémů jsou signály posloupnostmi a konvoluce je diskrétní, definovaná vztahem

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = x(n) * h(n). \quad (3.51)$$

Substitucí  $n - k = l$  do vztahu (3.51) lze dokázat komutativnost diskrétní konvoluce, čili vztah

$$x(n) * h(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = h(n) * x(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k). \quad (3.52)$$

Diskrétní konvoluci konečných signálů lze snadno nalézt bod po bodu graficky pomocí (3.51). Nejprve obrácením signálu  $h(k)$  získáme signál  $h(-k)$ . Opakováním posouváním tohoto signálu po jednom diskrétním kroku doprava a pro každý posuv vynásobením signálů  $x(k)$  a  $h(n-k)$  bod po bodu a sečtením takto získaných součinů získáme jednotlivé body konvoluce. Je-li délka signálů  $N_1$  a  $N_2$ , je délka signálu získaného konvolucí  $N_1 + N_2 - 1$ .

## 3.3 Diskrétní Fourierova transformace (DFT)

Frekvenční spektrum vzorkovaného signálu  $X(e^{j\theta})$  dle (2.22) je spojitá funkce normované kruhové frekvence  $\theta$  počítaná z nekonečného počtu hodnot  $x(n)$ . Při zpracování signálů pomocí číslicových obvodů se pracuje s konečnými počty hodnot. Pracuje se s konečnými posloupnostmi i ve frekvenční oblasti, tedy s konečnými počty vzorků frekvenčního spektra. Signály v oblasti času i frekvence mají stejný počet vzorků  $N$  a při výpočtech přímé i zpětné transformace se považují za periodické (pracujeme s periodickými prodlouženími signálů ze základního intervalu). Přechody mezi signály  $x(n)$  v časové oblasti a signály ve frekvenční oblasti, které budeme značit  $X(k)$ , je v tomto případě tzv. *finitní Fourierova transformace*. V případě diskrétních signálů se nazývá *diskrétní Fourierova transformace* a označuje se *DFT*. Tato transformace je definována vztahy

$$X(k/NT) = T \cdot \sum_{n=0}^{N-1} x(nT) e^{-j\frac{2\pi}{N}nk}, \quad k = 0, 1, 2, \dots, N-1, \quad (3.53)$$

$$x(nT) = \frac{1}{NT} \sum_{k=0}^{N-1} X(k/NT) e^{j\frac{2\pi}{N}nk}, \quad n = 0, 1, 2, \dots, N-1. \quad (3.54)$$

Rovnice (3.53) a (3.54) jsou rozměrově správné (hodnoty  $x(n)$  jsou ve [V], hodnoty  $X(k)$  ve [V/Hz]) a umožňují určit číselně frekvenci odpovídající dané hodnotě  $k$  ( $f_k = k/NT$ ) a čas odpovídající hodnotě  $n$  ( $t_n = nT$ ).

Při výpočtech DFT a IDFT se v praxi používají zjednodušené vztahy:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk}, \quad k = 0, 1, 2, \dots, N-1, \quad (3.55)$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}nk}, \quad n = 0, 1, 2, \dots, N-1. \quad (3.56)$$

Vztahy (3.53) a (3.55) se nazývají *přímá transformace*, vztahy (3.54) a (3.56) se nazývají *zpětná* resp. *inversní transformace*. Výjimečně se v literatuře vyskytuje koeficient  $1/NT$  resp.  $1/N$  v rovnicích pro přímou transformaci; pak není v rovnicích pro transformaci zpětnou. Přiřazení předmětu a obrazu v *DFT* budeme značit

$$x(n) \longleftrightarrow X(k). \quad (3.57)$$

DFT tedy vypočte  $N$  hodnot spektra  $X(k)$  z  $N$  hodnot signálu  $x(n)$ . Hodnoty spektra dostaneme pro diskrétní ekvidistantní hodnoty frekvencí začínající v  $f = 0$  a vzdálené od sebe o hodnoty

$$\Delta f = \frac{1}{NT} = \frac{f_{uz}}{N}. \quad (3.58)$$

Množině těchto hodnot se říká *DFT mřížka* (angl. *DFT grid*). Hodnota  $\Delta f$  se často v literatuře nazývá *frekvenční bin*.  $NT$  je celková doba odebírání signálu.

Pokud pracujeme s kruhovou frekvencí  $\omega$  resp. normovanou kruhovou frekvencí  $\theta$ , platí pro frekveční bin vztahy

$$\Delta\omega = \frac{2\pi}{NT} = \frac{\omega_{uz}}{N}, \quad (3.59)$$

$$\Delta\theta = \frac{2\pi}{N} = \Delta\omega \cdot T. \quad (3.60)$$

Vztah (3.55) plynne z (3.20), pokud změníme horní mez sumace a hodnoty  $X(e^{j\theta})$  počítáme v bodech  $\theta_k = k \cdot \Delta\theta = k \cdot (2\pi/N)$ .

**Příklad 3.1** Určete  $X(k)$  pro  $x(n)$  definované vztahy:  $x(n) = 2$  pro  $n = 4m + 1$ ,  $m$ -celé a  $x(n) = 0$  pro ostatní  $n$ ! ▲

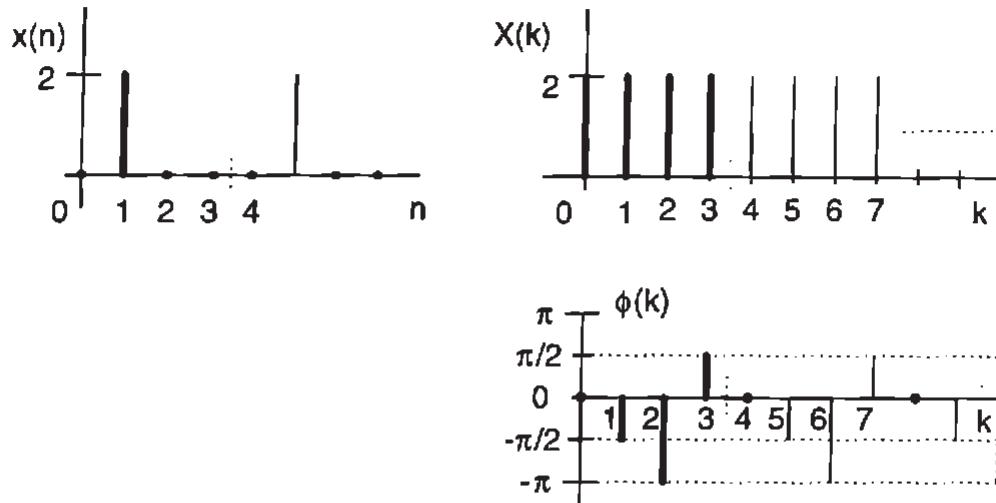
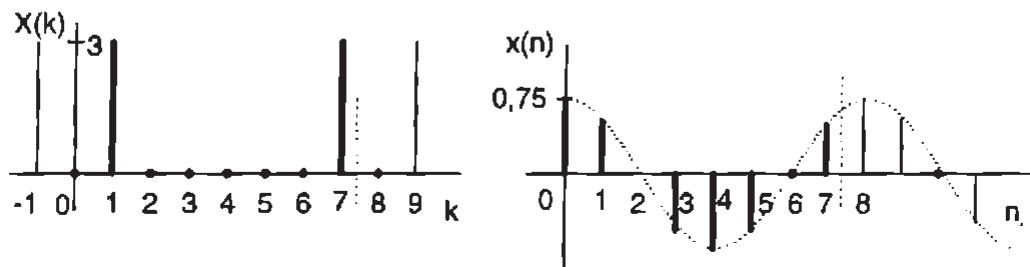
**Řešení** Dosazením  $N = 4$  do (3.55) dostaváme pro  $k = 0, 1, 2, 3$ :

$$X(k) = 2\delta(k) + 2e^{-j\pi/2} \cdot \delta(k-1) + 2e^{-j\pi} \cdot \delta(k-2) + 2e^{-j3\pi/2} \delta(k-3).$$

Graf signálu i jeho spektra uvádí obr. 3.8, kde svislá čárkovaná úsečka označuje konec základního intervalu. ■

**Příklad 3.2** Určete  $x(n)$  pomocí osmibodové IDFT, je-li

$$X(k) = 3\delta(k-1) + 3\delta(k-7), \quad 0 \leq k \leq 7.$$

Obrázek 3.8 Příklad čtyřbodové DFT. Základní interval je 0,1,2,3 pro  $n$  i  $k$ .

Obrázek 3.9 Spektrum a odpovídající signál k příkladu 2 (osmibodová IDFT).

**Řešení** Dosazením do (3.56) dostaneme:

$$x(n) = \frac{1}{8} \sum_{k=0}^7 X(k) e^{-j\frac{\pi}{4}kn} = \frac{3}{8} (e^{-j\frac{\pi}{4}n} + e^{-j\frac{\pi}{4}7n}) = \frac{3}{8} (e^{-j\frac{\pi}{4}n} + e^{j\frac{\pi}{4}n}) = \frac{3}{4} \cos\left(\frac{n\pi}{4}\right).$$

Graf signálu i jeho spektra uvádí obr. 3.9.

DFT spektra  $X(k)$  vypočtená dle (3.55) získáme vzorkováním FTD spekter vypočtených pomocí (3.20) v hodnotách  $\theta_k = (2\pi/N) \cdot k$ . Jde o vzorkování ve frekvenční oblasti, pro které platí obdobná vzorkovací věta, jako vzorkovací věta pro vzorkování v časové oblasti. Je-li tedy signál časově omezený (skládá se z  $N$  vzorků), může být pomocí vzorků ve frekvenční oblasti dokonale popsán.

### 3.3.1 Základní vlastnosti DFT

V dále uvedených rovnicích předpokládáme, že pracujeme s periodickými prodlouženými funkcí  $x(n)$  a  $X(k)$ , čili s funkcemi  $x(n+lN), X(k+lN)$ , kde  $k$  je celé. Tyto funkce dostaneme z (3.55) resp. (3.56), pokud se neomezíme na základní interval.

#### 1. Linearita

$$k_1 x_1(n) + k_2 x_2(n) = k_1 X_1(k) + k_2 X_2(k). \quad (3.61)$$

Vztah (3.61) lze zobecnit na lineární kombinace funkcí  $x_i(n)$  a  $X_i(k)$ . Pokud funkce  $x_i$  nejsou stejně dlouhé, doplní se kratší funkce na koncích nulami na délku nejdelší.

2. *Periodičnost*

Funkce  $x(n)$  i  $X(k)$  jsou periodické s periodou  $P = N$ .

3. *Kruhový časový posuv*

$$x(n - n_0) \longleftrightarrow e^{-j(\frac{2\pi}{N})kn_0} \cdot X(k), \quad n_0\text{-celé}. \quad (3.62)$$

Takto získaný posunutý průběh neodpovídá v intervalu  $(0, N - 1)$  průběhu, který bychom dostali posuvem funkce  $x(n)$  nulové mimo tento interval o  $n_0$  kroků. To je způsobeno periodičnosti funkce  $x(n)$ : posouváme periodické prodloužení signálu ze základního intervalu a vybereme část takto získaného průběhu, která padla do základního intervalu. V důsledku toho se hodnoty vysunované na jedné straně základního intervalu objevují na straně druhé.

4. *Kruhový frekvenční posuv*

$$e^{j(\frac{2\pi}{N})nk_0} x(n) \longleftrightarrow X(k - k_0), \quad k_0\text{-celé}. \quad (3.63)$$

Jde opět o kruhový posuv v základním intervalu popsáný v předchozím odstavci, ale tentokrát ve frekvenční oblasti.

Dvojím použitím vztahu (3.63) pro  $X(k - k_0)$  a  $X(k + k_0)$  dostaneme spektrum amplitudově modulovaného signálu  $2x(n) \cos((2\pi/N)nk_0)$ , protože kosinus můžeme zapsat pomocí exponenciál.

5. *Kruhová (cyklická) konvoluce v časové oblasti*

$$x_1(n) \circledast x_2(n) \longleftrightarrow X_1(k) \cdot X_2(k). \quad (3.64)$$

(Kruhová konvoluce je definována dále na str 29.)

6. *Kruhová (cyklická) konvoluce ve frekvenční oblasti*

$$x_1(n) \cdot x_2(n) \longleftrightarrow \frac{1}{N} X_1(k) \circledast X_2(k). \quad (3.65)$$

7. *Obraz obrácené posloupnosti*

$$x(-n) \longleftrightarrow X(-k). \quad (3.66)$$

Vztah (3.66) lze dokázat záměnou  $x(-n)$  za  $x(n)$  v sumě z (3.55), substitucí  $n = N - m$ , využitím vztahu  $X(m - N) = x(m)$  plynoucího z periodičnosti  $x(n)$  a využitím toho, že součet členů periodické posloupnosti přes celou periodu nezáleží na tom, od které hodnoty začneme sečítat.

Z (3.66) plynou další důležité závěry:

- $X(k)$  je sudá právě tehdy, je-li  $x(n)$  sudá.
- $X(k)$  je lichá právě tehdy, je-li  $x(n)$  lichá.

Je-li  $X(k)$  sudá, pak, protože je i periodická s periodou  $N$ , platí

$$X(k) = X(-k) = X(N - k). \quad (3.67)$$

Sudost posloupnosti má tedy za následek osovou symetrii hodnot  $x(n)$  a  $X(k)$  okolo středu základního intervalu  $(0, N - 1)$ .

8. *Vlastnosti spektra reálné posloupnosti*

Je-li  $x(n)$  reálná, pak je  $X(k)$  *konjugovaně sudá*, čili

$$X(k) = X^*(-k) = X^*(N - k). \quad (3.68)$$

Z (3.68) plynou následující vlastnosti spektra signálu, tvořeného posloupností reálných čísel (která v praxi získáme např. vzorkováním signálu ze snímače)

$$\operatorname{Re}[X(k)] = \operatorname{Re}[X(N-k)], \quad (3.69)$$

$$\operatorname{Im}[X(k)] = -\operatorname{Im}[X(N-k)], \quad (3.70)$$

$$|X(k)| = |X(N-k)|, \quad (3.71)$$

$$\phi(k) = -\phi(N-k). \quad (3.72)$$

Je-li  $x(n)$  reálná, stačí tedy znát hodnoty  $X(k)$  pouze pro  $0 \leq k \leq N/2$ .

9. *Vlastnosti spektra reálné a sudé posloupnosti*

Je-li  $x(n)$  reálná a sudá, pak i  $X(k)$  je reálná a sudá a platí

$$x(n) = x(N-n). \quad (3.73)$$

10. *Vlastnosti spektra reálné a liché posloupnosti*

Je-li  $x(n)$  reálná a lichá, pak  $X(k)$  je imaginární, lichá a platí

$$x(n) = -x(N-n). \quad (3.74)$$

11. *Alternativní vzorec pro výpočet IDFT*

Velký praktický význam má vztah

$$x(n) = \frac{1}{N} \cdot \left[ \sum_{k=0}^{N-1} X^*(k) e^{-j \frac{2\pi}{N} kn} \right]^*. \quad (3.75)$$

Vztah (3.75) plyne z (3.56), využijeme-li vztahů  $(X^*)^* = X$  a  $(e^{-j\phi})^* = e^{j\phi}$ . Rovnice (3.75) umožňuje vypočítat IDFT pomocí algoritmů (a programů) pro výpočet DFT. Při jeho použití nejprve obrátíme znaménka imaginárních částí hodnot  $X(k)$ , pak vypočteme DFT, změníme znaménka imaginárních částí vypočtených hodnot (čili vezmeme jejich komplexně sdružené hodnoty) a výsledek vydělíme  $N$ .

Kromě výše uvedených vlastností DFT lze při výpočtech DFT využít také možnosti rozkladu každé posloupnosti na součet liché a sudé posloupnosti dle (3.39). Často lze s výhodou využít také ortogonality exponenciálních funkcí, ze které plynou vztahy

$$\sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} nk} = \begin{cases} N & \text{pro } k = m \cdot N \text{ m-celé} \\ 0 & \text{pro ostatní k} \end{cases}, \quad (3.76)$$

resp.

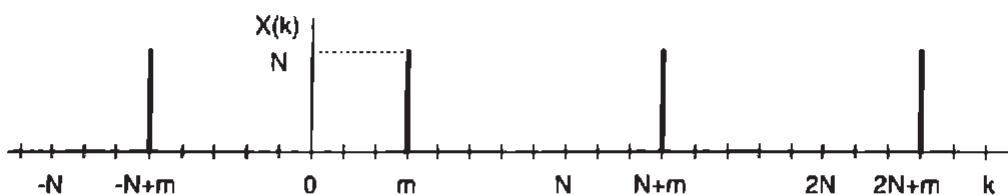
$$\sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} n(k-k_0)} = \begin{cases} N & \text{pro } k = k_0 + m \cdot N, m \text{- celé,} \\ 0 & \text{pro ostatní k.} \end{cases} \quad (3.77)$$

**Příklad 3.3** Vypočtěte spektrum komplexní exponenciály („komplexních kmitů“) s frekvencí  $f_0 = m\Delta f$ , resp.  $\theta_0 = m\Delta\theta = m \cdot (2\pi/N)$ , kde  $m$  je celé číslo. ▲

**Řešení**

$$x(n) = e^{jm(2\pi/N)n}, \quad (3.78)$$

$$X(k) = \sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} n(m-k)} = \begin{cases} N & \text{pro } k = m + lN, l \text{- celé,} \\ 0 & \text{pro ostatní m.} \end{cases} \quad (3.79)$$

Obrázek 3.10 Spektrum komplexní exponenciály dle (3.78) pro  $m = 3$ 

Spektrum je znázorněno v obr. 3.10. Odtud je zřejmé, že v grafech DFT spekter  $X(k)$  odpovídá hodnota  $N$  vzorkovací frekvenci, resp. hodnotě  $\theta_{vz} = 2\pi$ . V daném příkladě je frekvence komplexních kmitů  $f_0 = (3/8)f_{vz}$ . Porovnejte toto spektrum s FTD spektrem komplexní exponenciály (obr. 3.5).

### 3.3.2 Volba počtu bodů $N$ pro DFT

#### Volba počtu bodů DFT pro periodické diskrétní signály

Periodický diskrétní signál s periodou  $N_p$  je posloupnost, jejíž hodnoty se opakují vždy po  $N_p$  hodnotách, čili pro kterou platí

$$x(n + sN_p) = x(n), \quad s - \text{celé}. \quad (3.80)$$

Pro takové signály volíme

$$N = l \cdot N_p, \quad l - \text{přirozené}. \quad (3.81)$$

Při splnění (3.81) nedojde v periodickém pokračování  $x(n)$  na hranicích základního intervalu k nespojitostem obálky  $x(n)$  (analogové funkce  $x(t)$ , jejímž vzorkováním získáme  $x(n)$ ). Důsledkem toho je, že nedojde k *rozmažání spektra*, čili k prosakování energie ve spektru do sousedních frekvenčních binů, jevu zvanému anglicky *leakage*.

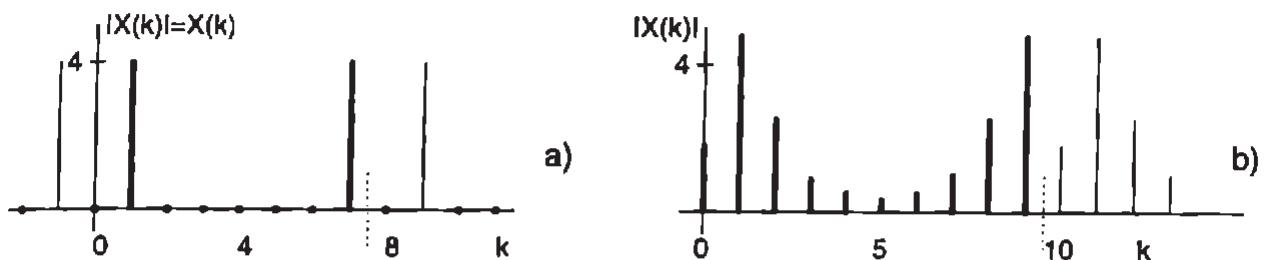
Leakage lze velmi názorně sledovat pro periodický diskrétní harmonický signál. Např. pro

$$x(n) = \cos \frac{2\pi n}{8} = \cos\left(\frac{\pi}{4}n\right) \quad (3.82)$$

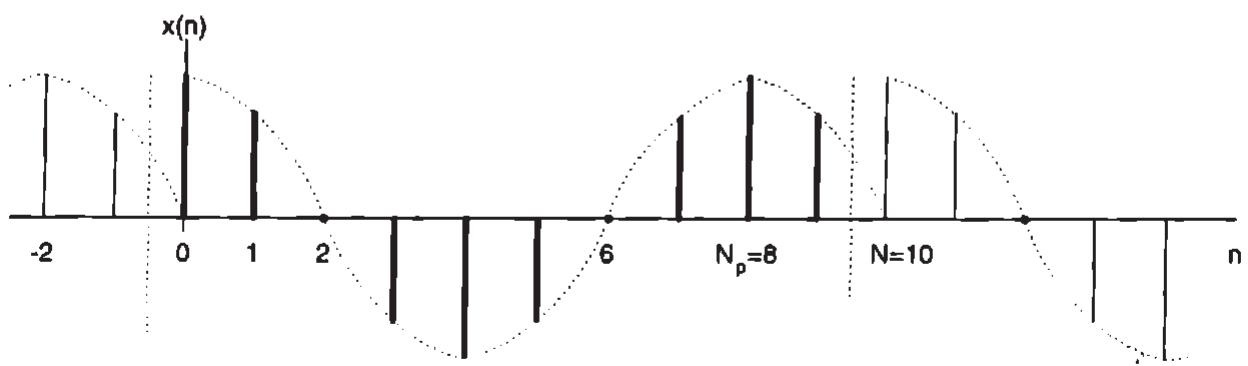
jde o signál s periodou  $N_p = 8$  a jeho spektrum je spektrem z obr. 3.9 vynásobeným  $4/3$ , pokud zvolíme  $N = N_p = 8$ . Uvádí je obr. 3.11a. Pokud ale spočítáme DFT pro  $N = 10$ , čili použijeme pro výpočet 10 bodů  $x(n)$ , bude spektrum komplexní (periodické prodloužení úseku  $x(n)$  v základním intervalu není sudá posloupnost) a amplitudové spektrum znázorňuje obr. 3.11b.

Pokud bychom chtěli v obr. 3.11 získat délku spektrální čáry odpovídající amplitudě kosinusovky, musíme spektrum normovat, čili vydělit  $N$  (pro jednostranné spektrum vydělit  $N/2$ ). Délka čáry pak bude 0,5, což odpovídá čáře kosinusovky s amplitudou 1 ve dvoustranném spektru.

V obr. 3.11a odpovídají čáry pro  $k = -1$  a  $k = 1$  dvoustrannému spektru kosinusovky, další čáry jsou důsledkem periodicity spektra. Čára pro  $k = 7$  odpovídá vztahu (3.71). Spektrum z obr. 3.11b odpovídá nespojitosti obálky  $x(n)$  v časové oblasti - viz obr. 3.12. Fourierova řada průběhu vzniklého periodickým prodloužením obálky takového signálu nemá pouze jeden člen.



Obrázek 3.11 DFT spektra periodické kosinusovky dle (3.82) ( $N_p = 8$ ) pro  $N = 8$  (a) a  $N = 10$  (b) - porušena podmínka (3.81)



Obrázek 3.12 Periodické prodloužení signálu dle (3.82) délky  $N = 10$

Druhou podmínkou, kterou se snažíme obvykle splnit při výpočtu DFT, je volba

$$N = 2^m, \quad m \text{-přirozené}. \quad (3.83)$$

Důvodem je skutečnost, že dnes se DFT počítá většinou pomocí algoritmů tzv. *rychlé Fourierovy transformace, FFT* (viz str. 31) a nejběžnější algoritmy FFT vyžadují splnění (3.83).

Pokud  $N_p \neq 2^m/l$ ,  $m, l$ -přirozené, nelze splnit současně (3.83) a (3.81).  $N$  se pak volí většinou dle (3.83) (u analyzátorů signálu ani jinou možnost většinou nemáme). Pro zmenšení chyby určení frekvencí jednotlivých složek spektra upravíme signál  $x(n)$  jeho *vynásobením časovým oknem*  $w(n)$  podle vztahu

$$x_w(n) = x(n) \cdot w(n). \quad (3.84)$$

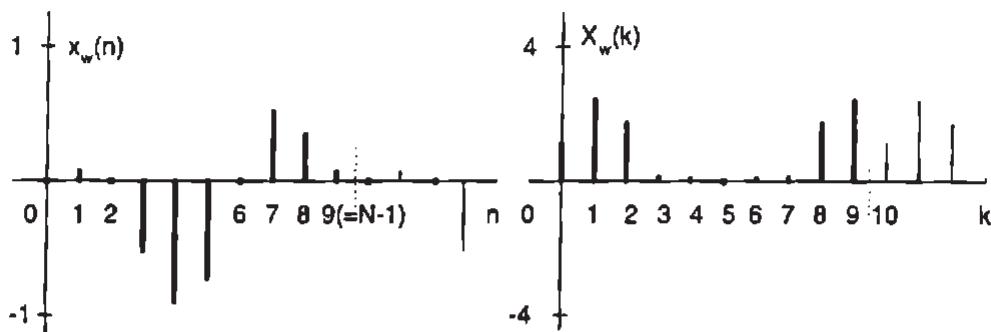
Téměř výlučně se používají okna s průběhem, který se i s několika derivacemi na obou koncích blíží nule. Použití takových oken zajistí mj. spojitost obálky periodického prodloužení signálu  $x_w(n)$ .

Casových oken (nazývaných též *datová okénka*) je celá řada. K nejrozšířenějším patří *okno Hann*, v literatuře (zejména americké) označované *hanning* resp. *Hanning*. Je definováno vztahem

$$w_{HN}(n) = \frac{1}{2}[1 - \cos(2\pi n/N)], \quad 0 \leq n < N. \quad (3.85)$$

Použijeme-li okno Hann na signál z obr. 3.12, dostaneme signál a spektrum z obr. 3.13.

Průběh spektra z obr. 3.13 je zřejmě bližší spektru z obr. 3.11a než spektrum z obr. 3.11b. Vynásobení signálu oknem má ovšem za následek změnu délek spektrálních čar. Frekvence



Obrázek 3.13 Průběh z obr. 3.13 upravený oknem Hann a jeho spektrum

odpovídající nejdelší čáře spektra také přesně neodpovídá frekvenci signálu. Obě tyto chyby lze odstranit v mnoha případech pomocí interpolace signálu ve frekvenční oblasti.

Protože v praxi analyzujeme vždy signál konečné délky, můžeme tento signál chápat jako nekonečně dlouhý signál vynásobený *obdélníkovým oknem* dle obr. 3.6a.

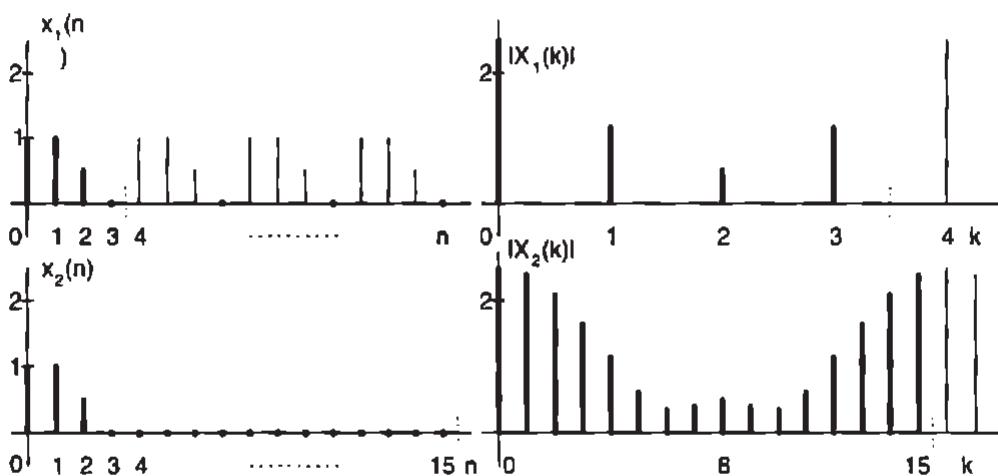
### Volba počtu bodů transformace $N$ pro signály konečné délky $N_s$

Pro tento případ volíme

$$N \geq N_s. \quad (3.86)$$

Volbou větší hodnoty  $N$  při zachování vzorkovací periody  $T$  zmenšujeme rozdíl frekvencí příslušný sousedním čárám ve spektru. Teoreticky je všechna informace obsažena v  $N_s$  vzorcích signálu, kterými je daný diskrétní signál konečné délky tvořen. Obálkou spektra je průběh  $X(e^{j\theta})$  získaný pomocí FTD (3.20). Zvyšováním hodnoty  $N$  pouze zvyšujeme počet vzorků spektra  $X(e^{j\theta})$  ve frekvenční oblasti vypočtených pomocí DFT.

Zvětšování  $N$  nad  $N_s$  znamená doplnění signálu na konci nulovými hodnotami (angl. *zero padding*). V praxi volíme  $N = 2^m$ . Doplnění signálu na konci nulami pomůže např. při detekci lokálních maxim spektra, jak to ukazuje obr. 3.14.

Obrázek 3.14 Získání mezilehlých hodnot ve spektru doplněním  $x(n)$  nulami ( $x_1(n) = x_2(n)$  pro  $0 \leq n < 4$ )

### Volba $N$ pro neperiodické časově neomezené diskrétní signály

Přesný popis signálů tohoto druhu je možný pouze pomocí FTD (3.20). V praxi je ovšem vždy počet vzorků konečný; je-li  $N$  dost velké, je chyba způsobená použitím DFT pro výpočet spektra zanedbatelná. Velikost této chyby závisí ovšem také na tvaru signálu. Maximální použitelná hodnota  $N$  je v praxi většinou omezena dostupnou pamětí počítače, případně programem pro DFT (resp. FFT, viz dále). Např. pro průběh amplitudového spektra signálu  $x(n) = (0,9)^n \cdot u(n)$  jsou chyby pro  $N = 64$  již v řádu 0,1% (viz [86], str. 74). Rozdíl hodnot  $X(e^{j\theta})$  a  $X(k)$  je způsoben useknutím zpracovávaného diskrétního signálu, čili jeho vynásobením *obdélníkovým (pravoúhlým) oknem* dle vztahu

$$w_0(n) = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & \text{jinde.} \end{cases} \quad (3.87)$$

### 3.3.3 Kruhová konvoluce a její užití

*Kruhová (cyklická) konvoluce* je definována vztahem

$$x_1(n) \circledast x_2(n) = \sum_{i=0}^{N-1} x_1(i) x_2(n-i) = \sum_{i=0}^{N-1} x_1(n-i) x_2(i). \quad (3.88)$$

Od *diskrétní aperiodické konvoluce* zvané též *lineární konvoluce* definované vztahem (3.51) se liší meziem sumace. Protože platí (3.64), používá se kruhová konvoluce zejména pro nalezení výstupu  $y(n)$  lineárního časově invariantního systému popsaného impulsní odevzrou  $h(n)$ , je-li na jeho vstupu signál  $x(n)$ :

$$y(n) = x(n) \circledast h(n). \quad (3.89)$$

Kruhová konvoluce je důležitá proto, že ji lze použít pro výpočet aperiodické konvoluce. Přitom alespoň jeden ze signálů v (3.51) musí být konečný (časově omezený). K výpočtu kruhové konvoluce lze využít platnosti vztahu (3.64) a počítat výstup pomocí vztahu

$$y(n) = IDFT\{X(k) \cdot H(k)\}. \quad (3.90)$$

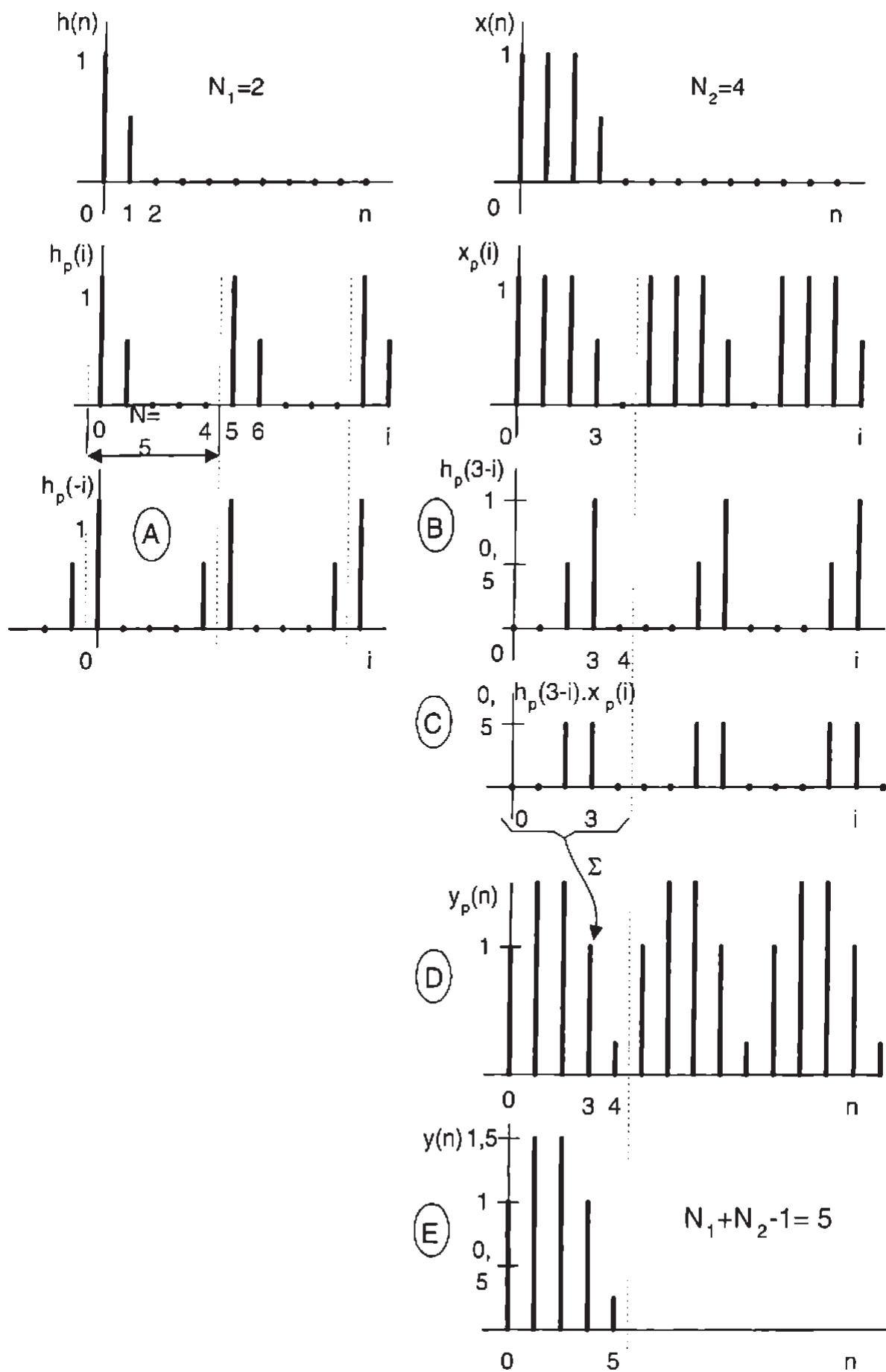
Výpočet podle (3.90) je rychlejší než podle (3.89), pokud pro výpočty  $X(k)$ ,  $H(k)$  a IFTD z (3.90) použijeme algoritmu rychlé Fourierovy transformace (str. 31).

### Výpočet aperiodické konvoluce dvou signálů konečné délky pomocí kruhové konvoluce

Chceme vypočítat (aperiodickou) konvoluci dvou signálů o délkách  $N_1$  a  $N_2$ . Víme (viz str. 21), že tato konvoluce bude mít délku  $N_1 + N_2 - 1$ . Tuto konvoluci můžeme dostat jako průběh cyklické konvoluce v základním intervalu délky  $N$ , zvolíme-li

$$N \geq N_1 + N_2 - 1. \quad (3.91)$$

Kruhová konvoluce je periodická s periodou  $N$ , což je současně délka DFT použité v (3.90). Proto před výpočtem doplníme oba signály nulami na délku  $N$  dle (3.91) a kruhovou konvoluci počítáme pro takto upravené signály. Pracujeme přitom s periodickým prodloužením signálů ze základního intervalu. Grafické nalezení kruhové konvoluce signálů  $h(n)$  a  $x(n)$  odpovídá vztahu (3.88): jeden ze signálů (ten kratší) převrátíme podle osy času (čili dostaneme signál  $h(-i)$ ). Hodnotu konvoluce pro  $n = 0$  určíme jako součet součinů hodnot signálu  $h(-i)$  a  $x(i)$  pro stejné  $i$ . Tento postup opakujeme  $n$ -krát, pro signál  $h(-i)$  posunovaný postupně



Obrázek 3.15 Výpočet aperiodické konvoluce pomocí konvoluce kruhové

o jeden krok doprava. Postup ze zřejmý z obr. 3.15, kde jsou jednotlivé kroky výpočtu podle vztahu (3.88) znázorněny velkými písmeny.

Pokud bychom pro výpočet kruhové konvoluce použili hodnotu  $N$  menší než odpovídá (3.91), dostali bychom průběh  $y(n)$  chybně.

### Použití DFT a kruhové konvoluce pro číslicovou filtraci

Filtrace signálu patří k základním operacím číslicového zpracování signálů. Základem číslicové filtrace je věnována samostatná kapitola. Protože se výstupní signál filtru vypočte jako konvoluce vstupního signálu a impulsní odezvy filtru, je možno k filtraci „nekonečně dlouhého“ signálu (např. výstupního signálu senzoru) filtrem s konečnou impulsní odezvou použít postupu popsaného v předchozím odstavci. Aperiodickou konvoluci vypočteme pomocí kruhové konvoluce a tuto pomocí rychlého výpočtu DFT (čili pomocí FFT).

Výpočet v časové oblasti podle vztahu (3.51) je pomalý a první hodnotu výstupu dostaneme až po skončení „nekonečně dlouhého“ signálu (v praxi ovšem vždy konečného, o délce  $N_2$ ). V daném případě platí  $N_2 \gg N_1$ , kde  $N_1$  je délka impulsní odezvy filtru. Není vhodný ani přímý výpočet podle vztahu (3.90). Řešením je *segmentace dlouhé posloupnosti*  $x(n)$  na řadu dílčích posloupností  $x_m(n)$  o délce  $L$  a výpočet výstupního signálu ze vztahu

$$y(n) = \sum_{i=-\infty}^{\infty} x_1(i)h(n-i) + \sum_{i=-\infty}^{\infty} x_2(i)h(n-i) + \dots = y_1(n) + \dots + y_m(n) + \dots \quad (3.92)$$

Tento vztah se nazývá *segmentovaná konvoluce*.

Posloupnosti  $y_i(n)$  jsou délky  $L + N_1 - 1$  a lze je počítat pomocí kruhové konvoluce při volbě  $N$  podle vztahu (3.91). Dílčí konvoluce spočteme najednou a sečteme je. Při sečítání ale musíme vzít v úvahu, že se dílčí signály  $y_i(N)$  částečně překrývají. Segmenty signálu  $x(n)$  mají délky  $L$  a navazují na sebe, zatímco dílčí konvoluce mají délky  $L + N_1 - 1$ . U každého segmentu se tedy překrývá  $N_1 - 1$  bodů, výstupy dílčích konvolucí začínají vždy současně s prvním bodem segmentu vstupního signálu. Popsaná metoda se nazývá *segmentovaná konvoluce s překryváním ve výstupní posloupnosti* (angl. *overlap-add method*).

Dílčí konvoluce se počítají pomocí *rychlé konvoluce* s využitím FFT a IFFT (dle vztahu (3.90)). Obraz kratší posloupnosti  $h(n)$  o délce  $N_1$  se pomocí FFT počítá jen jednou, a to po doplnění této posloupnosti nulami na délku  $N = N_1 + L - 1$ . Pak se použije FFT délky  $N$ . Nalezené hodnoty  $H(k)$  se opakovaně využijí při výpočtech  $Y_i(k)$  odpovídajících  $y_i(n)$  z rovnice (3.92).

Druhá použitelná metoda je *segmentovaná konvoluce s překryváním* ve vstupní posloupnosti. Při ní se posloupnost  $x(n)$  rozloží na dílčí posloupnosti, které se vzájemně překrývají. Více o této metodě viz např. ([68, 15]).

## 3.4 Rychlá Fourierova transformace (FFT)

Do šedesátých let bylo využití výpočtů DFT podle definice (vztahy (3.53) až (3.56)) omezeno na sálové počítače. Výpočty pro velké počty dat trvaly desítky minut i déle. Pro výpočet  $N$  hodnot pomocí (3.55) je třeba  $N^2$  komplexních násobení a  $N(N - 1)$  komplexních sčítání, takže doba potřebná pro výpočet je přibližně dána časem potřebným pro provedení  $N^2$  operací. S výjimkou speciálních obvodů pro číslicové zpracování signálu (tzv. signálových procesorů), kterým trvají všechny instrukce asembleru stejnou dobu, je pro celkovou dobu výpočtu rozhodující rychlosť násobení. Výpočet DFT pro  $N = 1024$  trval s prvními procesory INTEL desítky sekund,

s hardwarovou násobičkou desetiny sekundy. Dnešní signálové procesory pro tento výpočet potřebují čas okolo 1ms nebo kratší.

*FFT* čili *rychlá Fourierova transformace* (angl. *Fast Fourier transform*) jsou velmi efektivní algoritmy výpočtu DFT. Základní z nich popsali v r. 1965 J.W.Cooley a J.W.Tukey [8] a znamenal revoluci v číslicovém zpracování signálů. Ačkoliv podobné algoritmy byly popsány již začátkem 20. století, vzhledem k tehdejšímu stavu techniky nebyly prakticky využitelné [6, 7]. Dnes existuje algoritmu FFT celá řada. Základní z nich jsou popsány v [68, 5].

FFT má pro lineární zpracování signálů základní význam. Používá se pro spektrální analýzu deterministických i stochastických signálů, pro výpočet konvoluce, frekvenčních charakteristik i pro číslicovou filtrace.

Základní (a nejpoužívanější) algoritmy FFT jsou navrženy pro délku transformace  $N = 2^m$ , kde  $m$  je přirozené číslo. Tyto algoritmy využívají periodičnosti a symetrií komplexní exponenciály v (3.55). Exponenciála dle (3.93) se označuje *otáčecí činitel* (angl. *twiddle factor*) a označuje se  $W_N$ :

$$W_N = e^{-j\frac{2\pi}{N}}. \quad (3.93)$$

Pro tento činitel platí mimo jiné:

$$W_N^{kn} = W_N^{(k+N)n} = W_N^{k(n+N)}, \quad (3.94)$$

$$W_N^{2kn} = W_{N/2}^{kn}, \quad (3.95)$$

$$W_N^{k+n/2} = -W_N^k. \quad (3.96)$$

Algoritmy FFT pro  $N = 2^m$  redukují počet operací pro výpočet  $N$  bodů DFT na  $(N/2) \log_2 N = (Nm)/2$ . Poměrná úspora v počtu operací a tedy i době výpočtu exponenciálně roste s délkou transformace a výpočet pro  $N = 1024$  proběhne zhruba 200-krát rychleji (pro  $N = 64 \cdot 10^3$  zhruba  $10^4$ -krát rychleji) než při výpočtu podle definice DFT.

V současné době je některý z algoritmů FFT součástí většiny technických aplikačních programů. V dalším naznačíme odvození základního algoritmu FFT pro  $N = 2^m$ , algoritmu s *decimací v čase* (angl. *radix 2, decimation-in-time FFT algorithm*).

### FFT algoritmus s decimací v čase

S užitím vztahu (3.93) zapíšeme DFT jako

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk}, \quad k = 0, \dots, N-1. \quad (3.97)$$

Posloupnost  $x(n)$  rozdělíme na dvě dílčí posloupnosti, jednu obsahující pouze sudé členy [ $x_1(n) = x(2n)$ ] a druhou obsahující pouze liché členy [ $x_2(n) = x(2n+1)$ ], kde  $n \in \{0, N/2 - 1\}$ . Platí

$$\begin{aligned} X(k) &= \sum_{n=0}^{N/2-1} x(2n)W_N^{2nk} + \sum_{n=0}^{N/2-1} x(2n+1)W_N^{(2n+1)k} = \\ &= \sum_{n=0}^{N/2-1} x_1(n)W_N^{2nk} + W_N^k \sum_{n=0}^{N/2-1} x_2(n)W_N^{2nk}. \end{aligned} \quad (3.98)$$

Nyní využijeme vztahu (3.94) a dostaneme

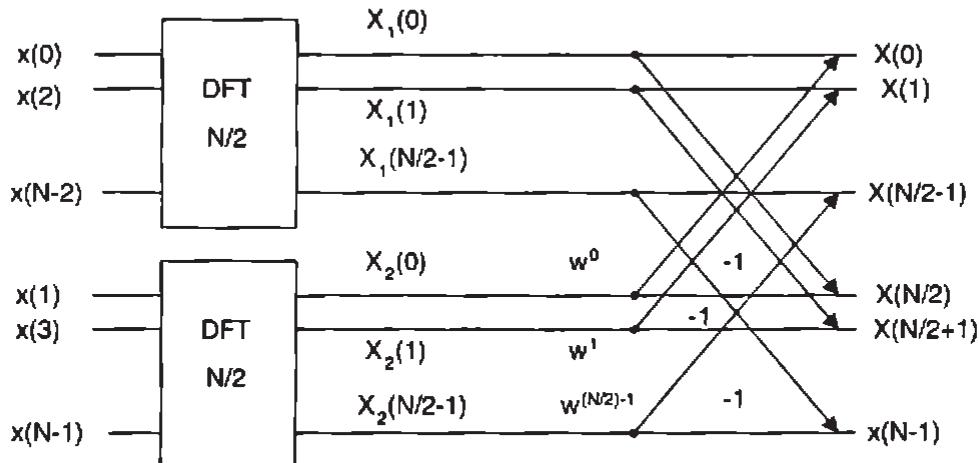
$$\begin{aligned} X(k) &= \sum_{n=0}^{N/2-1} x_1(n) W_{N/2}^{nk} + W_N^k \sum_{n=0}^{N/2-1} x_2(n) W_{N/2}^{nk} = \\ &= X_1(k) + W_N^k X_2(k), \quad k = 0, 1, \dots, N-1. \end{aligned} \quad (3.99)$$

Vztah (3.99) je součet dvou  $(N/2)$ -bodových DFT, aplikovaných na vybrané posloupnosti sudých a lichých členů signálu, přičemž  $X_2(k)$  je násobena mocninami  $W_N$ . Vzhledem k (3.94) a protože  $X_1(k)$  a  $X_2(k)$  jsou periodické s periodou  $N/2$  a  $W_N^k$  je periodická s periodou  $N$ , lze (3.99) rozepsat jako

$$\begin{aligned} X(k) &= X_1(k) + W_N^k X_2(k), \\ X(k+N/2) &= X_1(k) - W_N^k X_2(k), \end{aligned} \quad (3.100)$$

kde  $k = 0, 1, \dots, (N/2) - 1$ .

Vztahy (3.100) lze graficky znázornit podle obr. 3.16.



Obrázek 3.16 První krok algoritmu DIT FFT

Rozdělením původní  $N$ -bodové posloupnosti na dvě  $(N/2)$ -bodové posloupnosti ušetříme 50% operací. Místo  $N^2$  totiž potřebujeme nyní  $2 \cdot (N/2)^2 = N^2/2$  operací. Obdobné úspory dosahujeme dalšími obdobnými kroky, z nichž první naznačíme.

Rozdělením obou  $(N/2)$ -bodových posloupností stejným způsobem, jako jsme rozdělili signál, dostaneme čtyři  $(N/4)$ -bodové DFT:

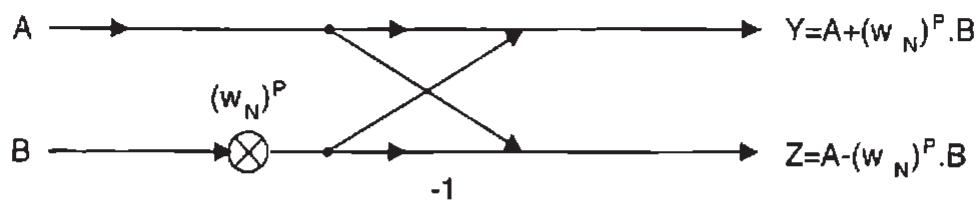
$$\begin{aligned} X_1(k) &= U(k) + W_N^{2k} V(k), \\ X_1(k + N/4) &= U(k) - W_N^{2k} V(k), \end{aligned} \quad (3.101)$$

$$\begin{aligned} X_2(k) &= Y(k) + W_N^{2k} Z(k), \\ X_2(k + N/4) &= Y(k) - W_N^{2k} Z(k), \end{aligned} \quad (3.102)$$

kde  $k = 0, 1, \dots, (N/4) - 1$ .

Vstupní posloupnosti  $U(k)$  a  $V(k)$  získáme pomocí DFT sudých a lichých členů  $x_1(n)$ , posloupnosti  $Y(k)$  a  $Z(k)$  získáme pomocí DFT sudých a lichých členů  $x_2(n)$ .

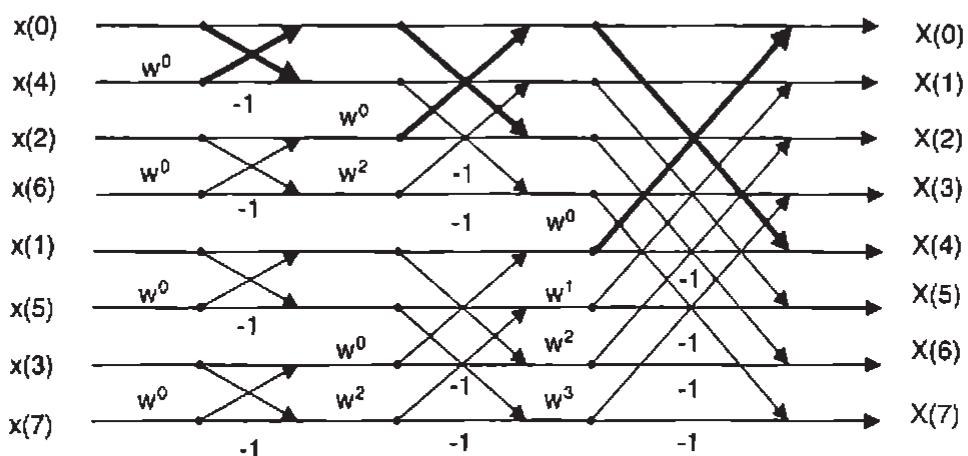
Opakováním tohoto postupu se dostaneme až k základní dvojici rovnic, která vyjadřuje dvoubodovou DFT. Znázornění těchto rovnic v grafu signálových toků (obr. 3.17) se podle



Obrázek 3.17 Motýlek DIT FFT

tvaru obázku nazývá *motýlek DFT*. Jedno vstupní číslo se násobí mocninou  $w_N^P$  a vypočítává se suma a rozdíl dvojice hodnot.

Čísla  $A$ ,  $B$ ,  $Y$ ,  $Z$  a  $w_N^P$  jsou v obecném případě komplexní. Komplexní exponenciálu  $w_N^P$  lze rozepsat pomocí sinu a kosinu a hodnoty těchto funkcí vybírat z tabulky. Celý výpočet FFT spočívá v kombinování motýlků podle systému, který pro osmibodovou DIT FFT znázorňuje obr. 3.18.



Obrázek 3.18 Schéma algoritmu osmibodové DIT FFT

Z obr. 3.18 je vidět, že vzdálenost vstupních bodů motýlků roste od dvoubodových transformací s mocninou 2 při přechodu do každého dalšího stupně algoritmu. Jednotlivé stupně se skládají ze skupin tvořených rostoucími počty motýlků.

Vstupní posloupnost v obr. 3.18 je přeházena způsobem, který získáme, zapíšeme-li indexy v pořadí 1 až 7 binárně a u každého provedeme *bitovou reverzaci*, čili přepíšeme bity původního čísla od konce k počátku. Např. pro  $3 = (011)_B$  dostaneme  $(110)_B = 6$ . U moderních signálových procesorů provádí bitovou reverzaci s minimální časovou ztrátou jeden z adresových generátorů.

Protože jsou počty vstupů a výstupů motýlků v jednotlivých stupních stejné, je možné přepsat původní vstupní signály výstupními. Tak se FFT zakončí ve stejné oblasti paměti, kterou zabírala vstupní data. Takové algoritmy se (anglicky) nazývají *in-place operation* algoritmy. Podrobný rozbor algoritmu DIT FFT i druhého základního algoritmu DIF FFT (algoritmus s decimací ve frekvenci) spolu s programem pro signálové procesory Analog Devices řady ADSP-21XX uvádí [39].

Při praktických výpočtech lze časovou úsporu získanou výše popsaným způsobem dále zvyšovat. Algoritmus DFT pracuje obecně pro komplexní vstupní posloupnost. Jsou-li vstupní

data  $x(n)$  reálná, lze vstupní posloupnost rozdělit na dvě poloviny ( $x(n) = x_1(n) + jx_2(n)$ ). Posloupnost  $x_1(n)$  se vloží jako reálné části vstupů a  $x_2(n)$  jako imaginární části vstupů. Pak se počítá  $(N/2)$ -bodová FFT. Po skončení výpočtu lze jednoduše oddělit obě poloviny výstupní posloupnosti [5, 30].

V současné době se pro výpočet FFT používají i jiné než základní algoritmy vyžadující  $2^m$  vstupních dat, např. algoritmy s komplikovanějšími motýlkami, vyžadující  $N = 4^m$  dat, případně algoritmy použitelné pro  $N$ , která jsou prvočíslky nebo součinem prvočísel. Podrobnou informaci o jiných možnostech výpočtu FFT lze najít např. v [68].

### 3.4.1 Vyjádření DFT pomocí matic

Protože otáčecí činitel (viz (3.93)) má jednotkovou velikost, lze výpočet DFT znázornit jednoduše maticovým schématem, kde je úhel otáčecího činitele naznačen směrem šipek jednotkových vektorů ve čtvercové matici. Posloupnosti  $X(k)$  a  $x(n)$  jsou sloupcové matice. Pro  $N = 8$  je výpočet DFT znázorněn na obr. 3.19

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \\ X(4) \\ X(5) \\ X(6) \\ X(7) \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow \\ \uparrow & \nearrow & \rightarrow & \nwarrow & \downarrow & \leftarrow & \leftarrow & \nwarrow \\ \uparrow & \rightarrow & \downarrow & \leftarrow & \uparrow & \rightarrow & \downarrow & \leftarrow \\ \uparrow & \nwarrow & \leftarrow & \nearrow & \downarrow & \nwarrow & \rightarrow & \nearrow \\ \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \uparrow & \leftarrow & \rightarrow & \nwarrow & \downarrow & \nearrow & \leftarrow & \nwarrow \\ \uparrow & \leftarrow & \downarrow & \rightarrow & \uparrow & \leftarrow & \downarrow & \rightarrow \\ \uparrow & \nwarrow & \leftarrow & \downarrow & \nwarrow & \rightarrow & \nearrow & \nearrow \end{bmatrix} \cdot \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \\ x(4) \\ x(5) \\ x(6) \\ x(7) \end{bmatrix}$$

Obrázek 3.19 Maticové schéma algoritmu osmibodové DFT; šipka vzhůru představuje reálnou jednotku

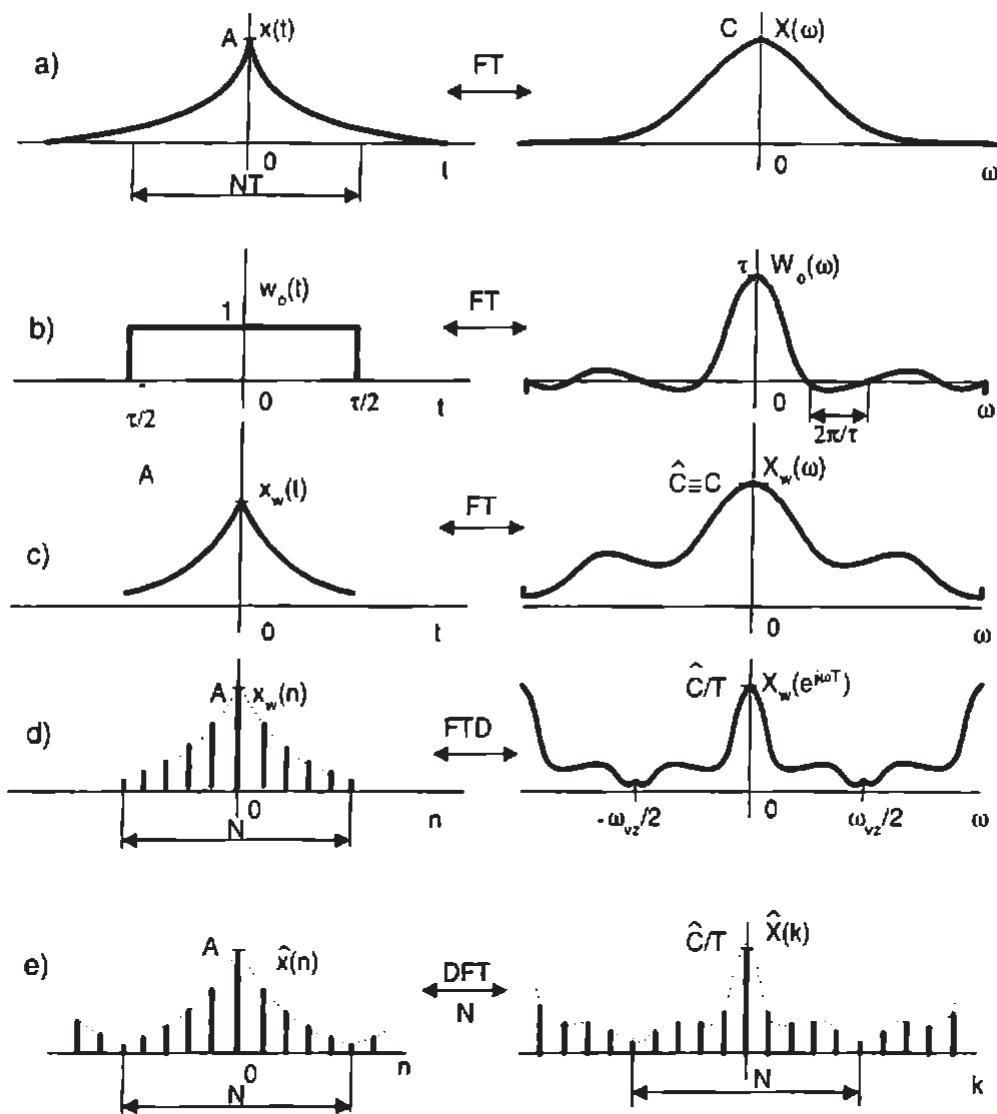
## 3.5 Použití DFT pro zpracování spojitéch signálů

DFT (a FFT) jsou definovány pro diskrétní signály. V praxi jsou ale velmi často používány pro přibližný výpočet spekter spojitéch signálů, často zjištěných měřením. Chyba určení spektra pak souvisí s tím, že se pracuje s konečnými počty hodnot v časové i frekvenční oblasti. Vliv jednotlivých kroků na výslednou souvislost původního spektra analogového signálu  $X(\omega)$  a spektra  $X(k)$  zjištěného DFT lze sledovat v obr. 3.20.

*Chyba useknutím* se projeví rozdílností spekter  $X_w(\omega)$  a  $X(\omega)$ , *aliasing* se projeví odchylkou spekter  $TX_w(e^{j\omega T})$  a  $X_w(\omega)$ . Obě tyto chyby zahrnuje rozdíl spekter  $TX(k)$  a  $X(\omega)$  na kruhových frekvencích  $\omega_k = k \cdot 2\pi f_{vz}/N$ .

Protože je nemožné, aby  $x(t)$  i  $X(\omega)$  byly pro týž signál konečné, vždy se projeví alespoň jedna z uvedených chyb. Chyba useknutím je nulová pro signály konečné délky, vezmeme-li dobu měření  $NT$  větší než je délka signálu. Chybu aliasing zmešujeme zvětšováním vzorkovací frekvence. To je ovšem provázeno rostoucími nároky na paměť a dobu výpočtu.

Uvedené chyby se projeví i v případě, že hodnoty vzorkovaného signálu mohou nabývat libovolné hodnoty v určitém intervalu (nedochází k zaokrouhlování při převodu na čísla) a že výpočty jsou prováděny s nulovou numerickou chybou. V praxi se ovšem vždy projeví šum analogově-číslicového převodu, chyby reálného AD převodníku, časová nejistota vzorkování



Obrázek 3.20 Vliv jednotlivých operací na změnu spektra analogového signálu v DFT spektru

(angl. *jitter*) - roztřesení vzorkovacích okamžiků, a konečná numerická přesnost použitého procesoru.

### 3.6 Využití DFT pro spektrální analýzu periodických signálů

Nalezení frekvenčního spektra signálu (frekvenční analýza) patří k důležitým úlohám měřicí a řídicí techniky. Používají se k ní přístroje zvané *spektrální analyzátory*, případně *analyzátory signálu*, které měří i další funkce. Spektrální analyzátory využívají buď filtrace signálu pomocí analogových nebo číslicových filtrů, nebo spektrum počítají pomocí DFT (v praxi FFT). K analýze lze využít i počítačů se zásuvnými deskami pro vzorkování signálu a vhodného programu.

FFT spektrální analýza umožňuje (na rozdíl od analýzy pomocí filtrů) určit amplitudové i fázové spektrum signálu a pomocí IFFT přejít zpět do časové oblasti. Operace FFT a IFFT se využívají i při měření korelačních a koherenčních funkcí a výkonových spektrálních hustot (viz str. 74).

DFT spektrum (a FFT spektrum) je vzorkovaným FTD spektrem, přičemž vzorky se odebírají v hodnotách  $\theta = k \frac{2\pi}{N}$ . V důsledku periodičnosti  $x(n)$  a  $X(k)$  u DFT se při analýze periodických signálů objeví problém, protože v praxi většinou nelze přesně splnit podmínu

$$f_{sig} \cdot T \cdot N = m = \frac{f_{sig}}{f_{vz}/N}, \quad m - \text{přirozené}. \quad (3.103)$$

Vztah (3.103) je pouze jinak zapsaná podmínka (3.81), neboť platí  $N_p = T_{sig}/T = f_{vz}/f_{sig}$ . Podmínu (3.103) lze v praxi splnit, pokud je vzorkování synchronizováno se signálem, což lze provést např. pomocí fázově vázané smyčky. Pak je na každou periodu signálu odcbráno přesně  $N$  vzorků a první z těchto vzorků je odebrán v okamžiku začátku periody signálu. Je-li podmínka (3.103) splněna, dostaneme pomocí FFT v polovině základního intervalu  $k \in (0, (N/2) - 1)$  odpovídající pásmu  $f \in (0, f_{vz}/2)$  nezkreslený průběh spektra. Toto spektrum se periodicky opakuje s periodou  $N$ , odpovídající hodnotě  $f_{vz}$ , resp.  $\omega_{vz}$ . Interval  $(0, f_{vz}/2)$  je zobrazen na stínítku FFT spektrálních analyzátorů. Toto pásmo je dostačující, protože měřené signály jsou posloupnostmi reálných čísel, pro které platí vlastnosti (3.71).

V praxi FFT spektrální analyzátory pracují s délkou transformace  $N = 1024$  (resp.  $N = 512$ ) a se vzorkovací frekvencí pevnou nebo nastavitelnou na malý počet hodnot, takže (3.103) není splněna. Pak dojde k *rozmažání spektra*, resp. *prosakování energie ve spektru do jiných frekvencí*, známějšímu pod anglickým označením *leakage*. Tento jev je pro kosinusový signál demonstrován v obr. 3.11. Zde ho budeme podrobněji analyzovat pro případ komplexní exponenciály:

$$x(n) = e^{j\theta_s n} = e^{j2\pi f_s T n}. \quad (3.104)$$

Spektrum takové funkce uvádí vztah (3.79) a obr. 3.10. Lze je zapsat také vztahem

$$X(k) = N \cdot \delta(k - m) \quad 0 \leq m \leq N - 1, \quad (3.105)$$

kde  $m$  je počet period exponenciály pokrytý právě  $N$  vzorky.

Pokud podmínka (3.103) není splněna, můžeme psát

$$f_{sig} \cdot N \cdot T = m + \alpha, \quad |\alpha| < 1/2, \quad m - \text{přirozené}. \quad (3.106)$$

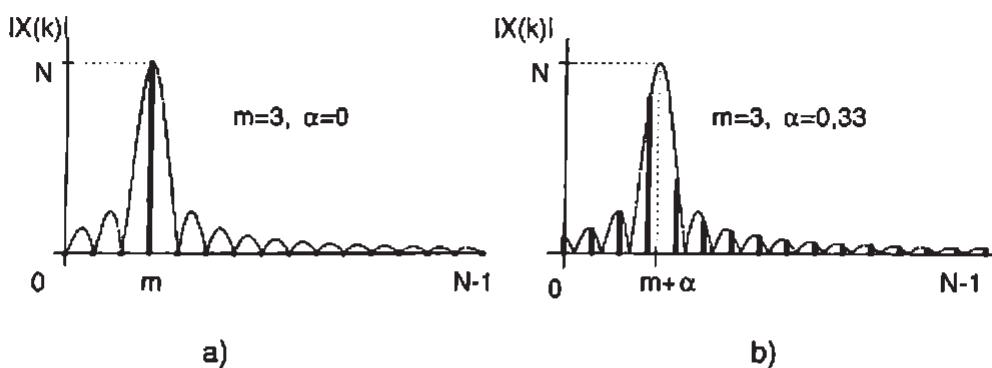
Pro spektrum exponenciály pak platí

$$X(k) = e^{j\pi(m-k+\alpha)\frac{N-1}{N}} \frac{\sin[\pi(m-k+\alpha)]}{\sin[\frac{\pi}{N}(m-k+\alpha)]} \quad (3.107)$$

a spektrum pak odpovídá obr. 3.21 b. V tomto případě není tedy spektrum (komplexní) exponenciály tvořeno v základním intervalu jedinou čarou, ale  $N$  čarami, které dostaneme jako vzorky spojitého FTD spektra obdélníkového okna s maximem posunutým do hodnoty  $m + \alpha$  a vzorkovaného v celých číslech od 0 do  $N - 1$ . Pro spektrum jsou nejdůležitější tři nejdelší čáry.

Přesnost FFT frekvenční analýzy lze zvýšit zvětšením intervalu měření  $T_m = N \cdot T$ , protože je vzdálenost sousedních čar spektra menší. Při dané vzorkovací frekvenci to znamená zvětšení  $N$ .

Obr. 3.21 a a 3.21b odpovídají případu analýzy konečného úseku signálu. To lze chápát jako analýzu nekonečného signálu násobeného obdélníkovým oknem délky  $N$ , viz (3.87). Spektrum tohoto signálu uvádí obr. 3.6. Amplitudová spektra obdélníkového impulsu jsou také průběhy v obr. 3.21. Nejvyšší oblouk se obvykle nazývá *hlavní lalok* nebo *hlavní oblouk*, ostatní oblouky jsou *postranní laloky*.

Obrázek 3.21 Spektrum komplexní exponenciály z (3.107) pro vyznačené hodnoty  $\alpha$ 

Pro frekvenční analýzu by zřejmě bylo nejvhodnější okno s co nejnižšími postranními lalky (potlačuje rušivé složky spektra na frekvencích vzdálenějších od frekvence signálu při nedodržení (3.103)) a s co nejužším hlavním lalokem s co největším poklesem amplitudy pro  $\alpha = 0,5$  (pak je ve spektru minimum velkých rušivých složek blízko frekvence signálu). Tyto požadavky bohužel současně nesplňuje žádné okno, proto se při DFT frekvenční analýze používají různá okna. DFT (resp. FFT) aplikujeme tedy ne na  $x(n)$ , ale na

$$x_w(n) = x(n) \cdot w(n). \quad (3.108)$$

Ve frekvenční oblasti odpovídá toto násobení konvoluci spekter  $X(e^{j\theta})$  a  $W(e^{j\theta})$ . Často používaným oknem je *okno Hann*, viz (3.85). Umístíme-li toto okno symetricky okolo počátku, bude popsáno rovnicí

$$w_{HN}(n) = \cos^2(\pi \frac{n}{N}) = \frac{1}{2} \left[ 1 + \cos(2\pi \frac{n}{N}) \right], \quad -\frac{N}{2} \leq n < \frac{N}{2} \quad (3.109)$$

Funkce (3.109) je sudá, takže má reálné spektrum. Po vyjádření kosinu pomocí exponenciál a využití vztahu (3.63) dostaneme FTD spektrum  $W_{HN}(e^{j\theta})$  jako superpozici tří posunutých spekter obdélníkového okna  $W_O(e^{j\theta})$  (viz obr. 3.22):

$$W_{HN}(e^{j\theta}) = \frac{1}{2} W_O(e^{j\theta}) + \frac{1}{4} W_O(e^{j(\theta - \frac{2\pi}{N})}) + \frac{1}{4} W_O(e^{j(\theta + \frac{2\pi}{N})}). \quad (3.110)$$

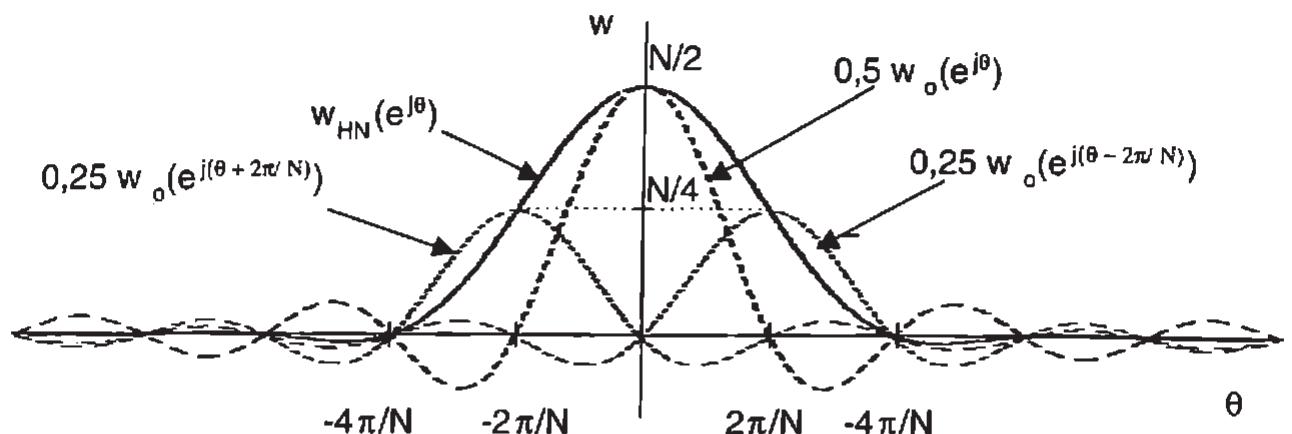
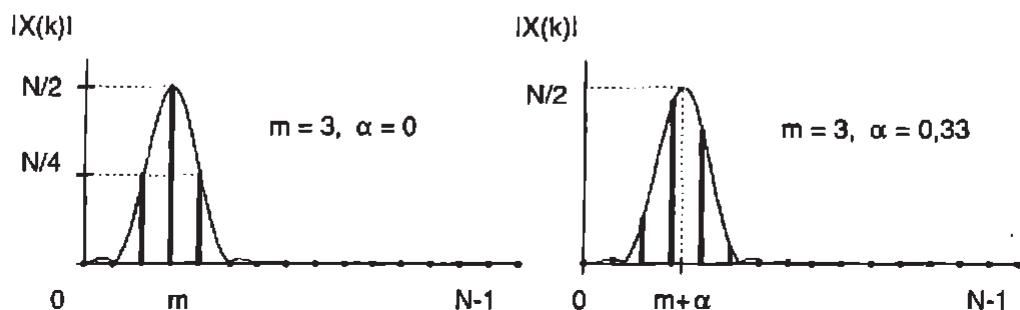
Z obr. 3.22 je zřejmé výrazné snížení postranních lalků proti spektru obdélníkového okna. Cenou za to je dvojnásobně širší hlavní lalok.

Použitím okna Hann na komplexní exponenciálu dostaneme podle splnění nebo nesplnění (3.103) spektra dle obr. 3.23.

Ze srovnání obr. 3.21 a obr. 3.23 je vidět, že pro  $\alpha = 0$  (splnění (3.103)) získáme spektrum bez chyby bez úpravy signálu (při užití obdélníkového okna). Neodpovídá-li  $N$  vztahu (3.103), potlačíme oknem Hann vzdálenější rušivé složky, ale objeví se místo jedné čáry tři, odpovídající vzorkování hlavního lalku z obr. 3.23 v bodech  $-2\pi/N, 0, 2\pi/N$ .

Další rozšířená okna (*Hamming, Blackman, Bartlett, Kaiser*) budou probrána v kapitole o číslicové filtrace, při které se také používají.

Vliv okna na průběh spektra lze demonstrovat při měření FFT spektra dvou kosinusovek blízkých frekvencí a podstatně odlišných amplitud (např.  $A_2 = 0.01A_1$ ). Lze li splnit podmínu (3.103) pro obě kosinusovky současně, je nejlepší obdélníkové okno příslušné délky. V opačném případě při rozmažání spektra silnějšího signálu bez použití vhodného okna slabší signál nezjistíme.

Obrázek 3.22 Konstrukce spektra okna Hann  $W_{HN}$  superpozicí tří spekter obdélníkového okna  $W_o$ .Obrázek 3.23 Spektrum komplexní exponenciály z (3.107) pro vyznačené hodnoty  $\alpha$  upravené oknem Hann

Některá časová okna se nemusejí aplikovat v časové oblasti (podle (3.108)), ale lze je snadno aplikovat ve frekvenční oblasti jako konvoluci spekter. U okna Hann stačí násobit tři po sobě následující vzorky spektra signálu konstantami  $-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}$ . Znaménka „-“ u  $\frac{1}{4}$  jsou důsledkem toho, že jde o okno podle (3.85), nikoliv podle (3.109).

Při spektrální analýze je výhodné mít možnost zkoumat vybraný úsek spektra podrobněji. Toto umožňují moderní FFT spektrální analyzátory v režimu *frekvenční lupy*. Menší vzdálenosti sousedních spektrálních čar vybraného úseku spektra se dosáhne pomocí *decimátoru*, který vybírá pouze každý  $m$ -tý vzorek ( $N$  se nemění). Vybraná část spektra se posune do okolí počátku násobením signálu komplexní exponenciálou (věta o posuvu v obraze). Komplexní exponenciál se ovšem skládá ze dvou harmonických signálů (podrobněji viz [30, 78]).

### 3.6.1 Souvislost DFT s koeficienty Fourierovy řady

Koeficienty komplexní Fourierovy řady periodické funkce  $x_p(t)$  jsou definovány vztahem

$$c_k = \frac{1}{T_0} \int_0^{T_0} x_p(t) e^{-j\omega_0 t} dt, \quad (3.111)$$

kde  $T_0$  je perioda funkce  $x_p(t)$  a  $\omega_0 = 2\pi/T_0$ .

Vzorkujeme-li funkci  $x_p(t)$   $N$  vzorky na periodu (čili platí-li  $NT = T_0$ ), lze vztah (3.111)

přibližně nahradit sumou (integrace pomocí obdélníkového pravidla)

$$c_k \doteq \frac{T}{NT} \sum_{n=0}^{N-1} x_p(nT) e^{-jk\omega_0 nT}. \quad (3.112)$$

Protože pro  $N$  vzorků na periodu je  $\omega_0 T = 2\pi T/(T_0) = 2\pi/N$ , platí

$$c_k \doteq \frac{1}{N} \sum_{n=0}^{N-1} x_p(nT) e^{-j\frac{2\pi}{N} kn}. \quad (3.113)$$

Označíme-li  $x_p(nT)$  jako  $x(n)$ , dostaneme s využitím (3.55) vztah

$$c_k \doteq \frac{1}{N} X(k), \quad (3.114)$$

kde  $X(k)$  je  $k$ -tá hodnota DFT spektra.

Vztah (3.114) je přibližný. Chyba náhrady integrálu sumou klesá s růstem  $N$  a poklesem  $T$  a závisí na průběhu signálu. Přesněji platí (viz [5])

$$\frac{1}{N} X(k) = c_k + c_{k-N} + c_{k+N} + c_{k-2N} + c_{k+2N} + \dots \quad (3.115)$$

Výraz  $(1/N)X(k)$  tedy vyjadřuje koeficient  $c_k$  Fourierovy řady v komplexním tvaru s chybou  $\Delta = \sum_{i=1}^{\infty} (c_{k-iN} + c_{k+iN})$ . Pokud hodnoty  $c_k$  klesají tak rychle, že je  $\Delta$  zanedbatelná proti  $c_k$ , platí vztah (3.114) prakticky přesně.

DFT (a FFT) je tedy použitelná pro harmonickou analýzu funkcí. Dává koeficienty Fourierovy řady v komplexním tvaru, ze kterých lze snadno vypočítat Fourierovy koeficienty  $a_k, b_k$  a  $A_k$ .

Jak plyne z (2.22) a z textu na str. 36, získáme pomocí DFT hodnoty spektra analogového signálu  $x(t)$  vzorkovaného s intervalom  $T$  na frekvencích  $\omega_k$ , případně  $\theta_k$  nebo  $f_k$  ze vztahů

$$X(j\omega_k) = X(jk \frac{2\pi}{N} f_{vz}) \doteq TX(k), \quad (3.116)$$

$$X(j\theta_k) = X(jk \frac{2\pi}{N}) \doteq TX(k), \quad (3.117)$$

$$X(jf_k) = X(jk \frac{1}{N} f_{vz}) \doteq TX(k). \quad (3.118)$$

## 4. Číslicové filtry

Návrh číslicových filtrů<sup>1</sup> tvoří spolu s DFT spektrální analýzou základ klasického číslicového zpracování signálů. Základní typy číslicových filtrů jsou lineární časově invariantní obvody s jedním vstupem a jedním výstupem. Lze je popsat přenosem (systémovou funkcí)  $H(z)$ , frekvenční charakteristikou (frekvenční odezvou)  $H(e^{j\theta})$ , impulsní odezvou (čili váhovou funkcí)  $h(n)$  nebo diferenční rovnicí.

Číslicová filtrace je nalezení výstupního signálu pro daný vstup a daný filtr. Výstup se získá jako (aperiodická) konvoluce vstupu a impulsní odezvy filtru. Při filtraci v reálném čase je následující vzorek výstupu odebrán až po výpočtu konvoluce, čili veškeré operace filtrace musí proběhnout v čase kratší, než je vzorkovací interval. Pro takovou filtraci jsou vhodné číslicové signálové procesory, jejichž všechny instrukce asembleru trvají stejně dlouho, a to řádově desítky  $ns$ . Jde o specializované procesory optimalizované z hlediska číslicového zpracování signálů.

Číslicový filtr je algoritmus nebo obvod, které požadovaným způsobem mění spektrum vstupního diskrétního signálu. Může být realizován speciálním obvodem nebo programem pro počítač. Ačkoliv původní oblastí jejich aplikace byla sdělovací technika, uplatňují se číslicové filtry dnes i v řadě jiných disciplín, mj. v měření a řízení. Tam jde např. o potlačení rušivých signálů superponovaných k signálům pracovním (potlačení signálů v určitých frekvenčních pásmech, různé typy průměrování), nebo o frekvenční analýzu signálů s využitím filtrace (pomocí přeladitelné pásmové propusti s úzkým propustným pásmem), využívanou v analýze vibrací v diagnostice elektromechanických soustav. Zajímavou aplikací je i dynamická korekce frekvenční odezvy senzorů (zvýšení mezní frekvence filtrací výstupu filtrem typu horní propust; cenou za to je pokles zesilení).

Číslicové filtry navazují na pasivní a aktivní analogové filtry a lze je navrhovat buď přímo, nebo převedením analogového prototypu do diskrétní oblasti (přechodem od přenosu  $H(p)$  k přenosu  $H(z)$ ).

Výhodou číslicových filtrů proti filtrům analogovým je možnost jejich využití pro nízké frekvence ( $f < 1Hz$ ), kde hodnoty  $L$  a  $C$  analogových filtrů a tedy i rozměry odpovídajících součástek vycházejí příliš velké, a zejména možnost změny vlastností číslicových filtrů změnou jejich koeficientů nebo změnou vzorkovací frekvence. Řád filtru je možno zvýšit opakováním přivedením výstupu filtru na jeho vstup. Změna vzorkovací frekvence mění mezní frekvenci filtru, změna koeficientů i typ filtru.

### 4.1 Rozdělení číslicových filtrů

Číslicové filtry se dělí podle délky impulsní odezvy na *filtry s konečnou impulsní odezvou*, obvykle označované *FIR filtry* (z angl. finite impulse response) a *filtry s nekonečnou impulsní*

<sup>1</sup>Protože se v této kapitole nebudeme až na malé výjimky zabývat problémy numerické přesnosti filtrů, měli bychom přesući mluvit o *diskrétních filtroch*. Vzhledem k praktické realizaci těchto filtrů pomocí číslicových obvodů používáme název číslicové filtry.

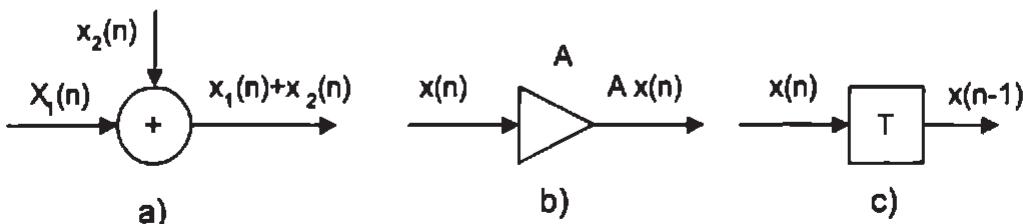
*odezvou*, obvykle označované *IIR filtry* (z angl. infinite impulse response). Z názvu přímo plynou podmínky pro počet členů impulsní odezvy  $h(n)$  obou typů filtru.

Podle struktury blokového schématu filtru se číslicové filtry dělí na *nerekurzívni filtry* (NRDF), které nemají zpětnou vazbu, a *rekurzívni filtry* (RDF), které obsahují zpětnovazební smyčku.

Nerekurzívni filtry jsou vždy FIR filtry. Rekurzívni filtry jsou téměř vždy IIR filtry, ale je možné je navrhnout tak, že jsou typu FIR.

Týž filtr lze realizovat různými obvody. Strukturu filtru se obvykle snažíme volit tak, aby pro zadaný přenos měl filtr minimální počet zpožďovacích členů. Pak jde o tzv. *kanonickou formu filtru*. Ta má pro impulsní odezvu  $h(n)$  délky  $N + 1$  ve své struktuře  $N$  zpožďovacích členů.

Číslicové filtry jsou lineární časově invariantní diskrétní systémy a jako takové je lze se staví ze tří základních bloků (viz obr. 4.1) - sčítáčky, násobičky (násobí vstup konstantou) a jednotkového zpoždění (zpožďuje signál o jeden vzorkovací interval  $T$ , resp. o hodnotu 1, pracujeme-li s argumenty signálů  $nT$ , resp.  $n$ ).



Obrázek 4.1 Základní bloky číslicových filtrů - sčítáčka (a), násobení konstantou (b) a jednotkové zpoždění (c)

#### 4.1.1 Číslicové filtry s konečnou impulsní odezvou - FIR filtry

Tyto filtry neobsahují zpětnou vazbu, jde tedy o filtry nerekurzívni. Je-li v nejdélší cestě mezi vstupem a výstupem filtru  $N$  zpožďovacích členů, má impulsní odezva  $N + 1$  členů.

Přenos  $H(z)$  kauzálního FIR filtru je dán vztahem

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_M z^{-M} = \sum_{n=0}^M b_n z^{-n}. \quad (4.1)$$

Impulsní odezva  $h(n)$  FIR filtru se získá zpětnou Z-transformací vztahu (4.1), a je tedy

$$h(n) = b_0 \delta(n) + b_1 \delta(n-1) + \cdots + b_M \delta(n-M). \quad (4.2)$$

Délka této odezvy (*délka filtru*) je tedy  $M + 1$ .

Vyjádříme-li  $H(z)$  v kladných mocninách  $z$ , dostaneme

$$H(z) = \frac{\sum_{n=0}^M b_n z^{M-n}}{z^M}. \quad (4.3)$$

Odtud je zřejmé, že  $H(z)$  má pól pouze v počátku (a to  $M$ -násobný). Filtry FIR jsou tedy vždy stabilní.

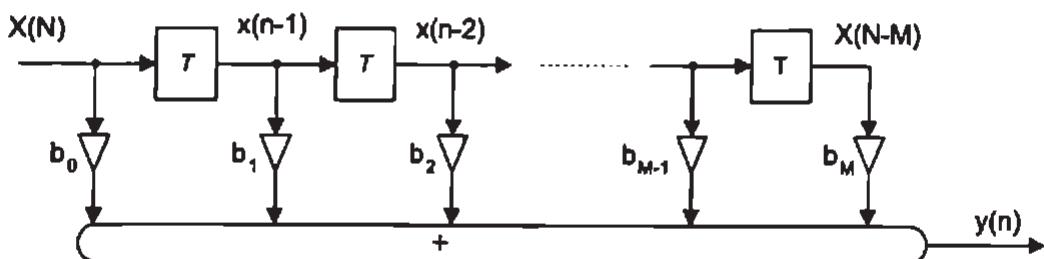
Z (4.1) lze odvodit *diferenční rovnici FIR filtru*:

$$y(n) = b_0 x(n) + b_1 x(n-1) + \cdots + b_M x(n-M). \quad (4.4)$$

Hodnota výstupního signálu závisí tedy na hodnotě vstupu a  $M$  předchozích hodnotách vstupu.

Nejrozšířenější strukturou FIR filtru je *transverzální filtr* (obr. 4.2). Jde o diskrétní zpožďovací linku s odbočkami, ve kterých jsou násobičky. Filtr počítá vážený klouzavý průměr  $M + 1$  posledních hodnot. Koefficienty násobiček jsou hodnotami impulsní odesady filtru (viz (4.2))

$$b_n = h(n) . \quad (4.5)$$



Obrázek 4.2 Blokové schéma transverzálního FIR filtru

Filtr zpožďuje vstupní signál o  $(M + 1)/2$  taktů, ustálený stav výstupu se dosáhne až po  $M + 1$  taktech. Řád FIR filtru  $M$  bývá v praxi vysoký (stovky až tisíce). Pak je totiž možno polynomem (4.1) velmi přesně approximovat požadovaný přenos. Nevýhodou je velké zpoždění výstupního signálu.

Důležitou vlastností FIR filtrů je možnost navrhovat je tak, aby měly *lineární fázi*. Fázová frekvenční charakteristika takového filtru je přímka se záporným sklonem, kterou lze vyjádřit vztahem

$$\varphi(e^{j\theta}) = -\alpha\theta + \beta, \quad \alpha, \beta > 0. \quad (4.6)$$

Častěji se pracuje se zápornou derivací fáze, zvanou *skupinové zpoždění*  $\tau_g(e^{j\theta})$ . Z (4.6) plyne

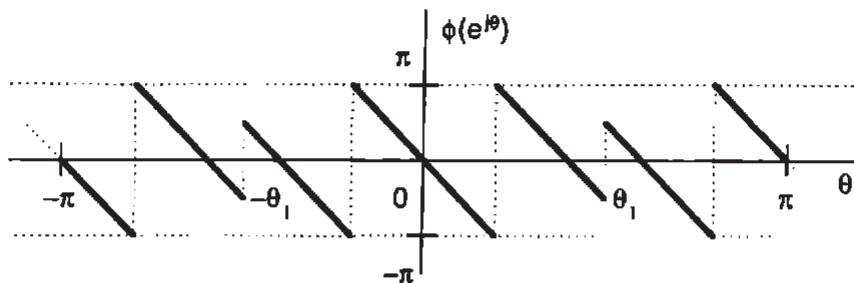
$$\tau_g(e^{j\theta}) = -\frac{d\varphi}{d\theta} = \alpha (= konst). \quad (4.7)$$

Filtry řádu  $M$  s lineární fází mají konstantní skupinové zpoždění rovné  $M/2$ , takže filtrovaný signál je zpožděn o  $M/2$  vzorkovacích kroků; jeho tvar není změněn, amplituda závisí na zesílení filtru. Tvar signálu v propustném pásmu filtru je zachován. (Pro filtr, který jen zpozdí signál bez změny jeho tvaru, platí  $y(n) = x(n - n_0)$  a podle věty o posuvu v předmětu DFT (vztah (3.27)) platí  $Y(e^{j\theta}) = e^{-jn_0\pi} \cdot X(e^{j\theta})$ . Přenos takového filtru je tedy  $H(e^{j\theta}) = 1 \cdot e^{-jn_0\pi}$  a jde tedy o filtr s lineární fází.)

Kromě zdánlivých skoků fáze z  $-\pi$  do  $\pi$  způsobených vynášením fázového posuvu v pásmu  $(-\pi, \pi)$  (viz poznámka pod čarou na str. 17) se v pile, ve kterou přejde přímka z (4.6), projeví navíc skoky fáze o  $\pi$ , odpovídající nulám  $H(z)$  ležícím na jednotkové kružnici. Proto může fázová charakteristika filtru s lineární fází mít např. tvar podle obr. 4.3. Skoky o  $\pi$  v hodnotách  $\theta_1$  a  $-\theta_1$  odpovídají nulám (v obrázku neuvedené) amplitudové frekvenční charakteristiky  $|H(e^{j\theta})|$ .

Frekvenční charakteristiku filtru najdeme dosazením  $z = e^{j\theta}$  do výrazu pro  $H(z)$ , tedy s užitím (4.1) a (4.5)

$$H(e^{j\theta}) = \sum_{n=0}^M h(n)e^{-jn\theta} = |H(e^{j\theta})| \cdot e^{j\varphi}. \quad (4.8)$$



Obrázek 4.3 Příklad fázové charakteristiky FIR filtru s lineární fází

Lze odvodit (viz [68]), že požadujeme-li průběh fázové charakteristiky dle (4.6), musí pro  $h(n)$  platit jeden z následujících vztahů

$$h(n) = h(M-n), \quad 0 \leq n \leq M, \quad \beta = 0, \quad \alpha = \frac{M}{2}, \quad (4.9)$$

$$h(n) = -h(M-n), \quad 0 \leq n \leq M, \quad \beta = \pm \frac{\pi}{2}, \quad \alpha = \frac{M}{2}. \quad (4.10)$$

Vztah (4.9) vyžaduje impulsní odezvu osové souměrnou podle osy procházející středem  $h(n)$  čili hodnotou  $\alpha$ , vztah (4.10) vyžaduje impulsní odezvu středově souměrnou podle středu  $h(n)$  čili hodnoty  $\alpha$ . Pro obě symetrie je nutno rozlišovat případy, kdy je řád filtru  $M$  sudé číslo a kdy je  $M$  liché číslo. Existují tedy 4 typy FIR filtrů s lineární fází, které ukazuje obr. 4.4. Tyto typy mají následující omezení průběhu frekvenční charakteristiky (viz např. [15]):

- typ 1 (sudá symetrie,  $M$  sudé): žádné omezení  $H(e^{j\theta})$ ,
- typ 2 (sudá symetrie,  $M$  liché):  $H(e^{j\pi}) = 0$ , filtr se nehodí pro návrh horní propusti,
- typ 3 (lichá symetrie,  $M$  sudé):  $H(e^{j0}) = 0$  a  $H(e^{j\pi}) = 0$ , filtr se hodí pouze pro návrh pásmové propusti,
- typ 4 (lichá symetrie,  $M$  liché):  $H(e^{j0}) = 0$ , filtr se nehodí pro návrh dolní propusti a pásmové zádrže.

Symetrie typu 1 a 2 se používají u základních frekvenčně selektivních filtrů (dolní a horní propust, pásmová propust a zádrž) (viz str. 49), symetrie typu 3 u derivátoru (str. 51) a typu 4 u Hilbertova transformátoru (str. 52).

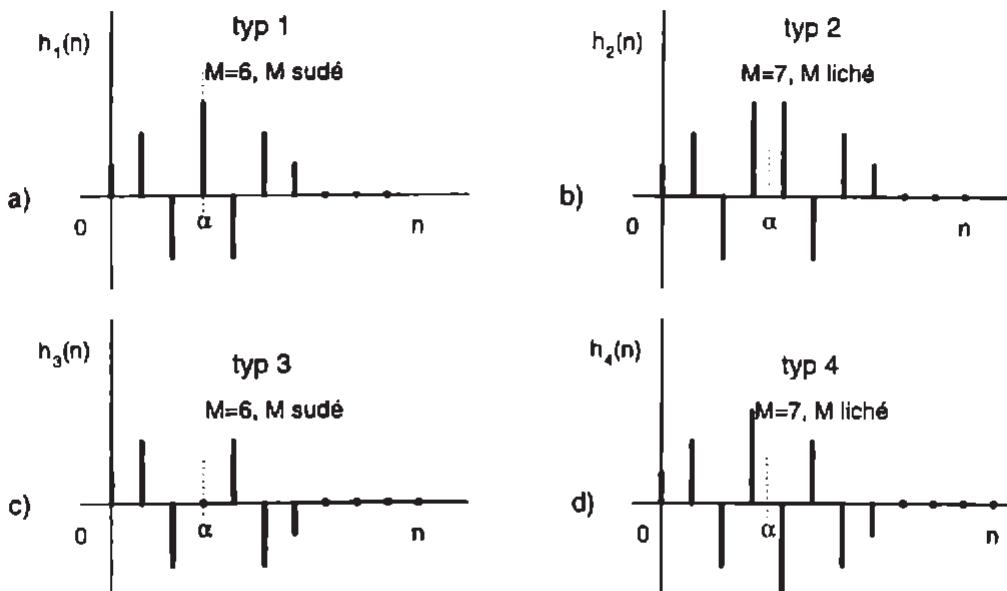
Pokud mají (nekauzální) FIR filtry impulsní odezvu sudou nebo lichou, mají fázi nulovou nebo rovnou  $\pi/2$ . Takové filtry se snadno navrhují. Filtry s lineární fází z nich získáme posuvem  $h(n)$  o  $L/2$  doprava. Tím se filtr stane kauzálním a v jeho přenosu se posuv projeví vynásobením původně reálného přenosu komplexní exponenciálou (věta o posuvu v předmětu FTD).

Využití uvedených symetrií pro snížení počtu násobení při návrhu filtru popisované v literatuře dnes již pozbyvá na významu, protože moderní signálové procesory potřebují pro násobení stejný čas jako pro ostatní instrukce.

FIR filtry s lineární fází se zadávají průběhem amplitudové frekvenční charakteristiky (případně s povoleným tolerančním pásmem okolo ní). Návrh spočívá v určení kauzální impulsní odezvy tak, aby byla zadaná funkce  $|H(e^{j\theta})|$  dostatečně přesně approximována a funkce  $h(n)$  vykazovala požadovanou symetrii.

#### 4.1.2 Číslicové filtry s nekonečnou impulsní odezvou - IIR filtry

Tyto filtry obsahují alespoň jednu zpětnovazební smyčku (s alespoň jedním zpožďovacím členem), jde tedy o rekurzivní číslicové filtry (RDF). Jejich přenos je tvořen podílem polynomů



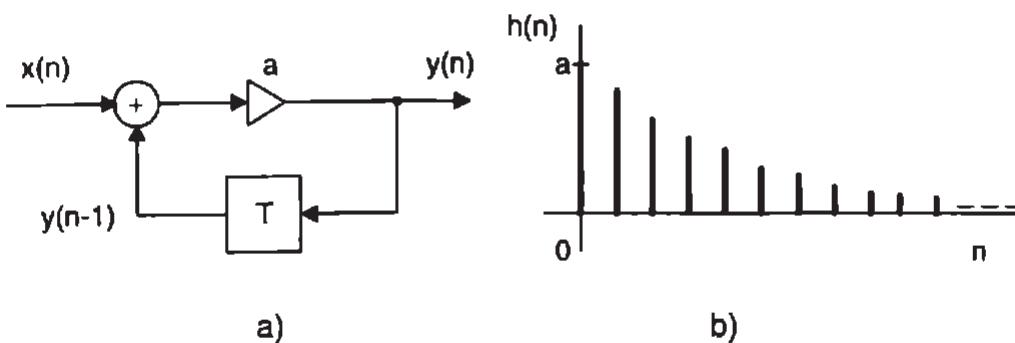
Obrázek 4.4 Čtyři typy impulsních charakteristik FIR filtrů s lineární fází

(racionální lomenou funkcí) a řád filtru je vyšší ze stupňů těchto polynomů vyjádřených v kladných mocninách  $z$ . Pro stejnou approximaci předepsané frekvenční charakteristiky mají IIR filtry podstatně nižší řád než FIR filtry, takže reagují rychleji na změny vstupního signálu. Mohou ale být nestabilní (mít póly vně jednotkového kruhu) a mají nelineární fázovou charakteristiku.

IIR filtr s daným přenosem lze realizovat dále uvedenými třemi způsoby: v přímé formě (první nebo druhé), v kaskádní formě a v paralelní formě. Z hlediska možných problémů se stabilitou způsobenou kvantováním koeficientů je nejvhodnější kaskádní forma, tvořená kaskádou (sériovým zapojením) filtrů druhého řádu.

Se zpětnou vazbou souvisí možnost vzniku *mezních cyklů* u IIR filtrů realizovaných v aritmetice s pevnou řádovou čárkou; RDF může za určitých okolností mít na výstupu periodický signál, i když vstup je nulový nebo konstantní. To je způsobeno zaokrouhlováním výstupu filtru použitého pro zpětnou vazbu v důsledku omezené numerické přesnosti výpočtu (více viz např. [68, 15, 86]).

Příkladem jednoduchého IIR filtru je např. obvod dle obr. 4.5. Jeho přenos je  $H(z) =$



Obrázek 4.5 Příklad IIR filtru (a) a jeho impulsní charakteristiky

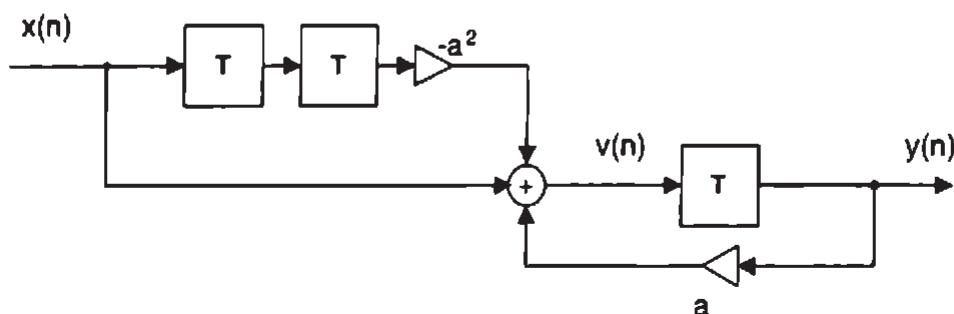
$(az)/(z - a)$  a jeho impulsní odezva je tedy  $h(n) = a \cdot a^n \cdot u(n) = a^{n+1} \cdot u(n)$ . Pro  $0 < a < 1$

znázorňuje  $h(n)$  obr. 4.5. Jde tedy o IIR filtr. Jeho nekonečná odezva odpovídá pólu  $p_1 = a$  mimo počátek.

RDF filtr může mít ve speciálním případě také konečnou impulsní odezvu. Příklad uvádí obr. 4.6. Tento obvod má přenos

$$H(z) = \frac{z^{-1} - a^2 z^{-3}}{1 - az^{-1}} = z^{-1} + az^{-2}.$$

Odtud zpětnou Z-transformací dostaneme  $h(n) = \delta(n-1) + a\delta(n-2)$ . Jde o speciální případ zpětné vazby, kdy v určitém taktu je součet všech vstupů do sčítátky na vstupu zpětnovazební smyčky umístěné na výstupu filtru nulový, takže v následujícím taktu impulsní odezva skončí. V praxi ale u obvodu s pevnou rádovou čárkou je přesné vyrušení vstupu do sčítátky v důsledku omezené numerické přesnosti koeficientů filtru obtížné.



Obrázek 4.6 Příklad rekurzivního filtru s konečnou impulsní odezvou

Rekurzivní číslicový filtr je popsán *přenosem* ve tvaru racionální lomené funkce

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_M z^{-M}}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_N z^{-N}}, \quad (4.11)$$

přičemž pro kauzální filtr je  $N \geq M$  a filtr popsáný (4.11) nazýváme filtr řádu  $N$ . Někdy je přenos IIR filtru vyjádřen ve tvaru blízkém (4.11), ale stupně obou polynomů jsou rovny  $N$ , přičemž několik posledních koeficientů  $b_i$  má nulové hodnoty.

Tvar přenosu uvedený v (4.11) (se zápornými znaménky ve jmenovateli a hodnotou  $a_0 = 1$ ) souvisí s tím, že (4.12) budeme používat ve tvaru<sup>2</sup>:

$$y(n) = \sum_{i=0}^M b_i x(n-i) + \sum_{i=1}^N a_i y(n-i). \quad (4.12)$$

První část rovnice (4.12) je vztah (4.4), čili diferenční rovnice FIR filtru. FIR filtry jsou tedy podskupinou IIR filtrů pro všechny koeficienty  $a_i, i \neq 0$ , nulové. Vzhledem k tomu, že se metody návrhu obou typů filtrů podstatně liší, probírají se obě skupiny odděleně.

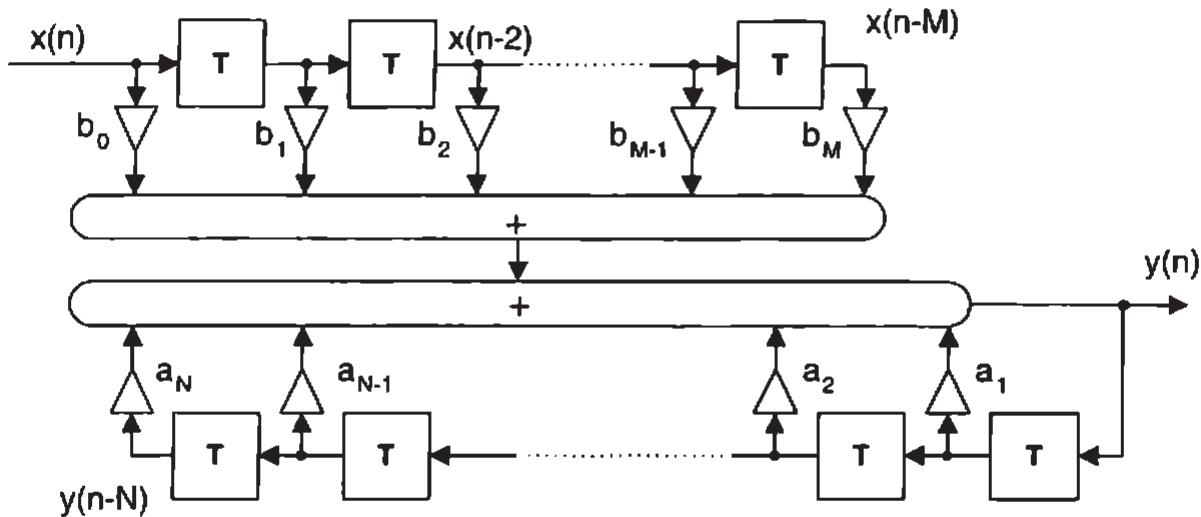
### První přímá forma realizace IIR filtru

První přímá forma realizace realizuje diferenční rovnici přímo ve tvaru (4.12) - viz obr. 4.7. Nuly přenosu realizuje nerekurzivní část schématu, póly rekurzivní část schématu. Jde o kaskádu

<sup>2</sup>V části literatury (a v programu MATLAB) se používá přenos s kladnými znaménky ve jmenovateli a  $a_0 \neq 1$ .

Na to je nutno dát pozor při používání MATLAB Signal Processing Toolbox.

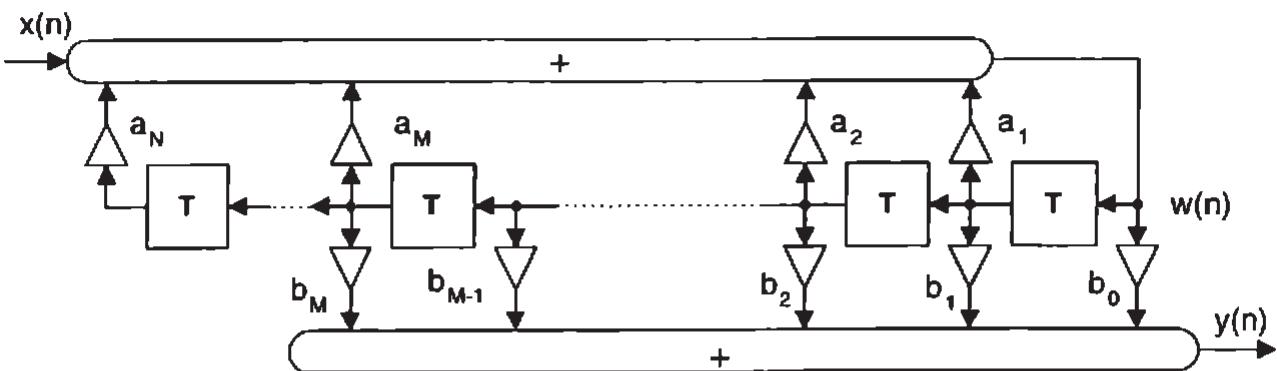
dvou diskrétních zpožďovacích linek s odbočkami. Nevýhodou první přímé formy je zbytečně velký počet zpožďovacích členů (jde o nekanonickou formu) a citlivost polohy všech nul a pólů na numerickou nepřesnost libovolného koeficientu.



Obrázek 4.7 Blokové schéma první přímé formy IIR filtru

### Druhá přímá forma realizace IIR filtru

Zaměníme-li pořadí zpožďovacích linek v první přímé formě, můžeme stejné zpožďovací členy použít pro obě linky. Místo  $M + N$  zpožďovacích členů pak potřebujeme pouze  $N$  těchto členů. Tato forma je kanonická a nazývá se *druhá přímá forma* (viz obr. 4.8). I když se vnitřní signály obou přímých forem liší, z hlediska vstupu a výstupu jsou obě formy ekvivalentní. Schéma z obr. 4.8 můžeme výhodně analyzovat s použitím pomocného signálu  $w(n)$ . Přenos



Obrázek 4.8 Blokové schéma druhé přímé formy IIR filtru

lze rozepsat jako

$$H(z) = \frac{Y(z)}{X(z)} = \frac{Y(z)}{W(z)} \cdot \frac{W(z)}{X(z)}. \quad (4.13)$$

Přitom platí

$$\frac{Y(z)}{W(z)} = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_M z^{-M}, \quad (4.14)$$

$$\frac{W(z)}{X(z)} = \frac{1}{1 - a_1z^{-1} - a_2z^{-2} - \dots - a_Nz^{-N}} \quad (4.15)$$

Diferenční (rekurzivní) rovnice odpovídající dílcům přenosům dle (4.14) a (4.15) jsou

$$y(n) = b_0w(n) + b_1w(n-1) + \dots + b_Mw(n-M), \quad (4.16)$$

$$w(n) = x(n) + a_1w(n-1) + \dots + a_Nw(n-N). \quad (4.17)$$

Tento zápis je výhodný pro filtraci pomocí signálového procesoru. Stačí si pamatovat signál  $w(n)$  v jediném kruhovém registru, předpokládá se  $M = N$  a koeficienty  $b_{M+1}, b_{M+2}, \dots, b_N$  se položí rovné nule.

I druhá přímá forma je citlivá na numerickou nepřesnost koeficientů - změna jediného z nich (způsobená numerickou nepřesností procesoru) ovlivní polohu všech nul a pólů přenosu.

### Kaskádní forma realizace IIR filtru

Tato struktura odpovídá rozepsání  $H(z)$  IIR filtru do součinu racionálních lomených funkcí 2. stupně (bikvadratických členů). Je-li řád filtru lichý, je koeficient u kvadratického členu jednoho z kvadratických dvojčlenů roven nule. Přenos této struktury je

$$H(z) = \frac{\sum_{i=0}^M b_i z^{-i}}{1 - \sum_{i=1}^N a_i z^{-i}} = b_0 \prod_{k=1}^P \frac{1 + B_{1k}z^{-1} + B_{2k}z^{-2}}{1 - A_{1k}z^{-1} - A_{2k}z^{-2}} = b_0 \prod_{k=1}^P H_k(z), \quad (4.18)$$

kde  $P = N/2$  pro  $N$  sudé a  $P = (N+1)/2$  pro  $N$  liché. Protože  $B_{1k}, B_{2k}, A_{1k}$  a  $A_{2k}$  jsou reálná čísla, jsou nuly i póly bikvadratických členů reálné nebo komplexně sdružené dvojice. Dvojčleny v čitateli i jmenovateli bikvadratického členu lze rozepsat do součinu kořenových činetelů v  $z^{-1}$ , čili jako  $(1 - z_{k1}z^{-1})(1 - z_{k2}z^{-1})$ . V případě komplexně sdružených kořenů je  $z_{k2} = z_{k1}^*$ . Je-li  $M < N$ , je příslušný počet hodnot  $B_{1k}$  a  $B_{2k}$  nulový.

Blokové schéma odpovídající vztahu (4.18) uvádí obr. 4.9.



Obrázek 4.9 Blokové schéma kaskádní formy IIR filtru

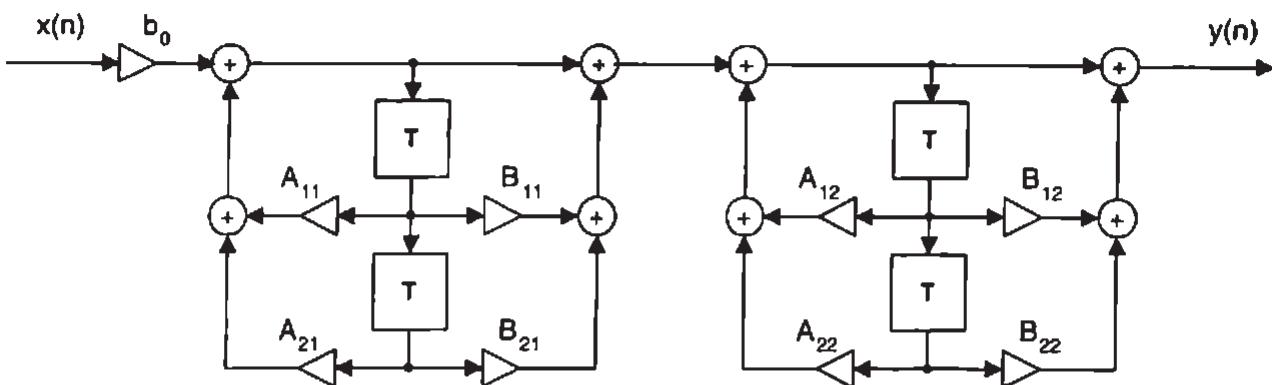
Jednotlivé bloky  $H_k(z)$  se realizují pomocí druhé přímé formy a pro  $N = 4$  uvádí příslušné blokové schéma obr. 4.10.

Rozdělení nul a pólů  $H(z)$  do jednotlivých sekcí záleží na navrhovateli filtru. V důsledku konečné numerické přesnosti jsou některé kombinace výhodnější. Výhodou kaskádové struktury, která je v praxi nejpoužívanější, je to, že nepřesnost určitého koeficientu ovlivní pouze polohu nejvýše dvou nul nebo pólů jediné sekce  $H(z)$ .

### Paralelní forma realizace IIR filtru

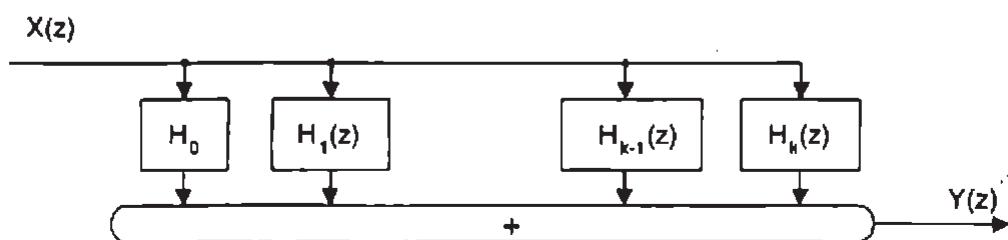
Paralelní forma je náhradou  $H(z)$  z výrazu (4.11) součtem parciálních zlomků  $H_i(z)$ .

$$H(z) = H_0 + H_1(z) + H_2(z) + \dots + H_k(z), \quad (4.19)$$



Obrázek 4.10 Kaskádní forma IIR filtru 4. stupně

kde  $H_0$  je konstanta a  $H_i(z)$  jsou bikvadratické sekce jako v (4.18), ve kterých opět některé hodnoty  $A_i$  a  $B_i$  mohou být nulové. Odpovídající blokové schéma uvádí obr. 4.11.



Obrázek 4.11 Paralelní forma IIR filtru

### 4.1.3 Rozdělení číslicových filtrů podle účelu

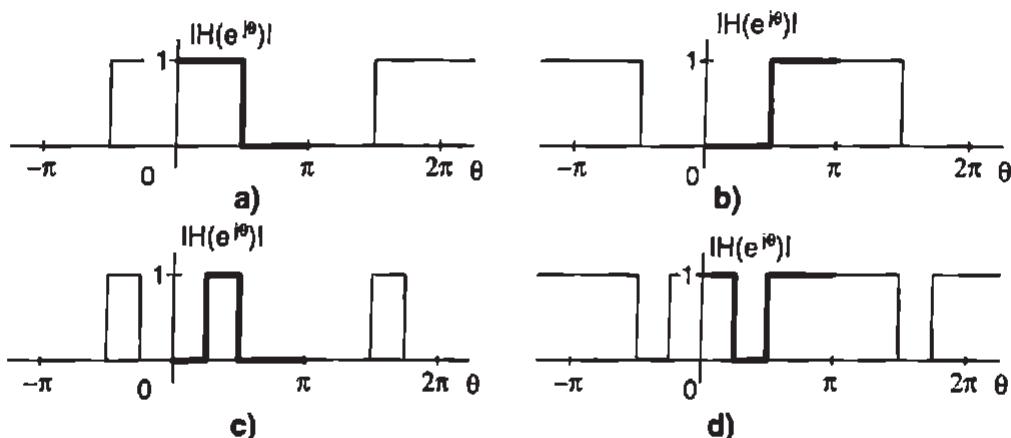
V předchozí části bylo probráno rozdělení číslicových filtrů podle délky impulsní odezvy a způsobu vyjádření přenosu filtru. Nyní se zmíníme o třech důležitých skupinách filtrů podle účelu filtrů. Jak uvidíme, v některých případech je názornější popis filtrace ve frekvenční oblasti, v jiných případech v časové oblasti.

Amplitudové frekvenční charakteristiky číslicových filtrů jsou (na rozdíl od charakteristik analogových filtrů) periodické s periodou  $f_{vz}$ ,  $\omega_{vz}$  resp.  $2\pi$  podle toho, vynášíme-li charakteristiku jako funkci  $f$ ,  $\omega$  nebo  $\theta$ . Protože impulsní odezvy filtrů jsou reálné posloupnosti, jsou amplitudové frekvenční charakteristiky sudé funkce a fázové frekvenční charakteristiky liché funkce. Stačí tedy znát jejich průběh v polovině základního intervalu. Vzhledem k požadavku vyloučení aliasingu je tvar charakteristiky v polovině základního intervalu určující pro typ filtru.

#### Frekvenčně selektivní číslicové filtry

Jde o filtry, jejichž účelem je propustit nebo potlačit složky signálu v určitém frekvenčním pásmu. Základní typy těchto filtrů pracují s jediným (souvislým) propustným nebo závěrným pásmem. Tyto filtry se používají zejména k potlačení periodických nebo šumových rušivých signálů superponovaných k signálům pracovním. Tak se zvyšuje poměr užitečného signálu k rušení.

Základními typy frekvenčně selektivních filtrů jsou *dolní propust* (DP), *horní propust* (HP), *pásmove propust* (PP) a *pásmove zádrž* (PZ) (obr. 4.12). Název filtru odpovídá průběhu amplitudové frekvenční charakteristiky v polovině základního intervalu.



Obrázek 4.12 Idealizované amplitudové frekvenční charakteristiky frekvenčně selektivních filtrů: dolní propust (a), horní propust (b), pásmová propust (c), pásmová zádrž (d)

Filtry tohoto typu se realizují jako IIR filtry i FIR filtry. Návrh vychází z návrhu dolní propusti, ze které se frekvenční transformacemi přejde k jinému typu filtru. Fázová frekvenční charakteristika se nezadává; pokud se požaduje lineární, použije se FIR filtr vhodného typu z obr. 4.4.

### Další důležité typy diskrétních filtrů

V řídicí a měřicí technice jsou důležité také další skupiny číslicových filtrů - integrátory, derivátory, fázové posouvače a filtry provádějící Hilbertovu transformaci a vypočítávající klouzavý průměr. U nich je většinou vyžadovaná úprava signálu vyjádřena primárně v časové oblasti. Pro realizaci algoritmů se využívá zejména metod známých z numerické matematiky.

### Diskrétní integrátor

Přenos ideálního analogového (spojitého) integrátoru v Laplaceově transformaci je  $H_{Iid}(p)$ <sup>3</sup> a jeho frekvenční charakteristika je  $H_{Iid}(\omega)$ :

$$H_{Iid}(p) = \frac{1}{p}, \quad H_{Iid}(e^{j\omega}) = \frac{1}{j\omega}. \quad (4.20)$$

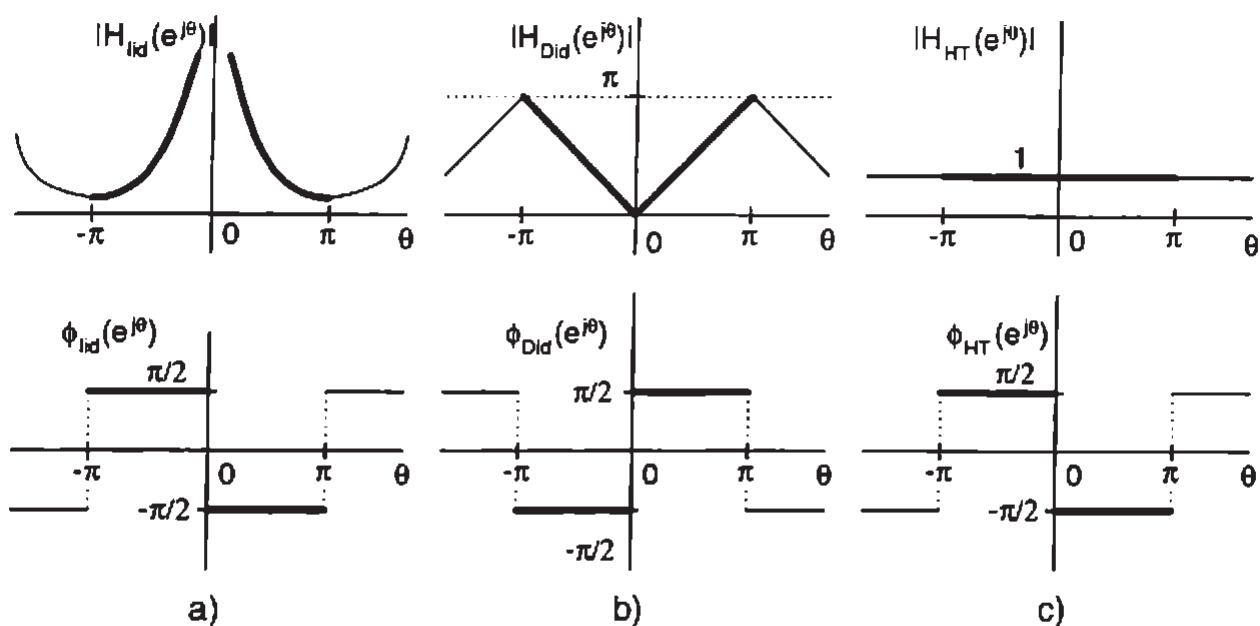
Odpovídající frekvenční charakteristika ideálního diskrétního integrátoru je

$$H_{id}(e^{j\theta}) = \frac{1}{j\theta} = \frac{1}{\theta} e^{-j\frac{\pi}{2}}. \quad (4.21)$$

Znázorňuje ji obr. 4.13 a.

Diskrétní integrátory ve skutečnosti počítají integrál pomocí některého z algoritmů známých z numerické matematiky. Nejjednodušší jsou následující rekurentní (diferenční) rovnice:

<sup>3</sup>Komplexní proměnná v Laplaceově transformaci se často označuje  $s$ , my budeme používat výše uvedené označení  $p$ .



Obrázek 4.13 Frekvenční charakteristiky ideálních diskrétních filtrů - integrátoru (a), derivátoru (b) a Hilber-tova transformátoru (c)

(4.22) pro obdélníkové pravidlo, (4.23) pro lichoběžníkové pravidlo, (4.24) pro Simpsonovo pravidlo a (4.25) pro Ticksovo pravidlo (viz [85]).

$$y(n) = y(n-1) + x(n), \quad (4.22)$$

$$y(n) = y(n-1) + \frac{1}{2}[x(n) + x(n-1)], \quad (4.23)$$

$$y(n) = y(n-2) + \frac{1}{3}x(n) + \frac{4}{3}x(n-1) + \frac{1}{3}x(n-2), \quad (4.24)$$

$$y(n) = y(n-2) + 0,3584x(n) + 1,2832x(n-1) + 0,3584x(n-2). \quad (4.25)$$

Je-li signál  $x(n)$  posloupnost vzorků napětí odebraných v ekvidistantních okamžicích, musíme hodnoty  $x(n)$  vynásobit vzorkovacím intervalom  $T$ . Hodnoty integrálu pak vyjdou správně ve [Vs]. Bloková schémata, přenosy, frekvenční charakteristiky a chyby uvedených integrátorů uvádí např. [86]. Integrátor Simpsonův integruje s malou chybou pro  $\theta \in (0, \pi/2)$ , čili pro  $f \in (0, f_{uz}/4)$ . Širší pásmo integrace s malou chybou vykazuje integrátor Ticksův. Vysokofrekvenční signály (s frekvencemi blízko  $f_{uz}/2$ ) vyhlaďuje lepě integrátor obdélníkový a lichoběžníkový.

### Diskrétní derivátor

Filtr derivující vstupní signál je důležitý např. pro určování zrychlení z rychlosti u časových průběhů nebo pro detekci hran při zpracování obrazů. Jeho ideální frekvenční charakteristika odpovídá spojité frekvenční charakteristice  $H_{Did}(\omega) = j\omega$  a je

$$H_{Did}(e^{j\theta}) = j\theta, \quad |\theta| \leq \pi. \quad (4.26)$$

Tuto charakteristiku uvádí obr. 4.13 b.

Nejjednodušším přibližným výpočtem derivace je náhrada derivace diferenci. Podle druhu použité diference dostaneme vztah

$$y(n) = x(n) - x(n-1), \quad (4.27)$$

nebo s využitím rozdílu dopředné a zpětné diference (nekauzální) filtr popsaný rekurentním vztahem

$$y(n) = \frac{1}{2}[x(n+1) - x(n-1)]. \quad (4.28)$$

Dokonalejší derivování umožňuje nekauzální filtr popsaný diferenční rovnicí ([85, 45])

$$y(n) = \frac{1}{12}[x(n-2) - 8x(n-1) + 8x(n+1) - x(n+2)]. \quad (4.29)$$

Podstatně dokonalejší derivátor můžeme navrhnout dále popsanou metodou FTD a oken pro návrh FIR filtrů (str.59). Chceme-li navrhnout kauzální derivátor rádu  $M$ , musíme impulsní odezvu  $h(n)$  zvolenou jako lichou posloupnost posunout o  $M/2$  kroků doprava. Ideální frekvenční charakteristika kauzálního derivátoru je

$$H_{DidK}(e^{j\theta}) = j\theta \cdot e^{-j\alpha\theta}, \quad |\theta| \leq \pi, \quad (4.30)$$

kde  $\alpha = M/2$ . Odpovídající impulsní odezvu dostaneme postupem uvedeným na str. 59 a platí

$$h_{DidK}(n) = \frac{(-1)^{n-\alpha}}{n-\alpha}, \quad n \leq M, \quad \alpha = M/2. \quad (4.31)$$

Zlepšení průběhu (potlačení zvlnění) frekvenční charakteristiky kauzálního derivátoru podle vztahu (4.31) se dosáhne vynásobením  $h_{DidK}(n)$  vhodným oknem stejné délky jako je  $h(n)$  (např. oknem Hann). Přitom approximace ideální charakteristiky je tím lepší, čím vyšší je rád filtru  $M$ .

Přenosy a frekvenční charakteristiky těchto derivátorů<sup>4</sup> lze nalézt např. v [86].

### Hilbertův transformátor

Další důležitý FIR filtr je *Hilbertův transformátor*. Jde o obvod, který se používá pro získání imaginární složky tzv. analytického signálu (viz [68]) ze složky reálné. Jeho ideální charakteristiky uvádí obr. 4.13 c, odkud je zřejmé, že jde o ideální fázový posouvač posouvající fázi v základním intervalu o  $-\pi/2$  pro  $\theta > 0$  a o  $+\pi/2$  pro  $\theta < 0$ . Jeho ideální frekvenční charakteristika je

$$H_{HT}(e^{j\theta}) = \begin{cases} -j, & 0 \leq \theta < \pi, \\ j, & -\pi \leq \theta < 0. \end{cases} \quad (4.32)$$

Reálná a imaginární část Fourierova obrazu kauzální posloupnosti (tj. posloupnosti nulové pro  $n < 0$ ) jsou spolu totiž jednoznačně spojeny, a to vztahem nazývaným Hilbertova transformace. Ta se využívá v řadě aplikací zpracování signálu. Zde se o ní zmiňujeme pouze pro úplnost a zájemce odkazujeme např. na [68, 43].

<sup>4</sup>Přibližně je možno derivátor realizovat také jako kaskádní zapojení ideálního derivátoru a dolní propusti. Tento obvod se někdy nazývá „pseudoderivátor“ a jde o IIR filtr získaný transformací analogového filtru s přenosem

$$H_D(\omega) = \frac{j\omega}{1 + a j\omega}.$$

## Fázový posouvač

Fázové posouvače jsou skupinou obvodů, které mají jednotkovou amplitudovou frekvenční charakteristiku (takže neovlivňují amplitudu signálu), ale mění průběh fázové charakteristiky na požadovaný. Používají se v kombinaci s jinými filtry, aby bylo dosaženo žádaného průběhu výsledné fázové charakteristiky (např. ve spojení s IIR filtry pro dosažení přibližně lineárního průběhu fáze v určitém frekvenčním intervalu). Hilbertův transformátor dle (4.32) je tedy příkladem fázového posouvače. Zvláštním příkladem je obvod s reálným jednotkovým přenosem, zvaný anglicky *all-pass*. Má význam pro některé teoretické úvahy (viz str. 65).

## Klouzavý průměr

Filtr realizující klouzavý průměr počítá průměr z daného vzorku a  $M$  vzorků předchozích. Jde tedy o FIR filtr s transverzální strukturou dle obr. 4.2, pro který  $\sum_{i=0}^M b_i = 1$ . Nejdůležitější je tzv. *nevážený klouzavý průměr*, jehož všechny koeficienty jsou stejné a platí pro ně  $b_i = 1/(M+1)$ ,  $i \in \{0, M\}$ . Kromě něj je v praxi důležitý také *klouzavý průměr s exponenciálním zapomínáním*, kde koeficienty mají váhy  $a^{-ci}$ , kde  $c$  je přirozené číslo. Protože bývá zvykem mluvit o klouzavém průměru délky  $N$ , budeme dál označovat délku tohoto filtru  $N$  namísto běžně používaného označení délky FIR filtrů  $L$  (platí pro ni  $L = M + 1$ ). Anglicky se tento filtr nazývá *moving average* a označuje se zkratkou  $MA$ .

Tento filtr se používá v technice např. pro potlačení periodického rušení superponovaného na konstantní signál, v ekonomice pro vyhlazování trendu (potlačení krátkodobých výkyvů časové řady).

*Nevážený klouzavý průměr* je popsán diferenční rovnicí

$$\begin{aligned} y(n) &= \frac{1}{N} \sum_{i=n-(N-1)}^n x(i) = \frac{1}{N} [x(n - (N - 1)) + \\ &+ x(n + 1 - (N - 1)) + \dots + x(n - 1) + x(n)]. \end{aligned} \quad (4.33)$$

Přenos filtru je tedy

$$H_{MA}(z) = \frac{1}{N} \sum_{i=0}^{N-1} z^{-i}. \quad (4.34)$$

Vztah (4.34) je ovšem součet  $N$  členů geometrické řady s kvocientem  $z^{-1}$ , takže lze psát

$$H_{MA}(z) = \frac{Y(z)}{X(z)} = \frac{1}{N} \frac{1 - z^{-N}}{1 - z^{-1}}. \quad (4.35)$$

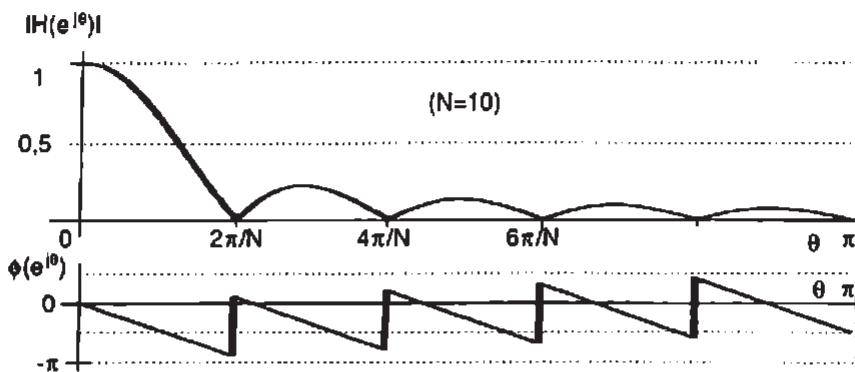
Odpovídající frekvenční charakteristika je tedy

$$H_{MA}(e^{j\theta}) = H_{MA}(z)|_{z=e^{j\theta}} = \frac{1}{N} \frac{1 - e^{-jN\theta}}{1 - e^{-j\theta}} = \frac{1}{N} e^{-j(N-1)\frac{\theta}{2}} \frac{\sin(N\frac{\theta}{2})}{\sin\frac{\theta}{2}}. \quad (4.36)$$

Amplitudová a fázová frekvenční charakteristika tedy jsou

$$|H_{MA}(e^{j\theta})| = \frac{1}{N} \frac{\sin(N\frac{\theta}{2})}{\sin\frac{\theta}{2}}, \quad (4.37)$$

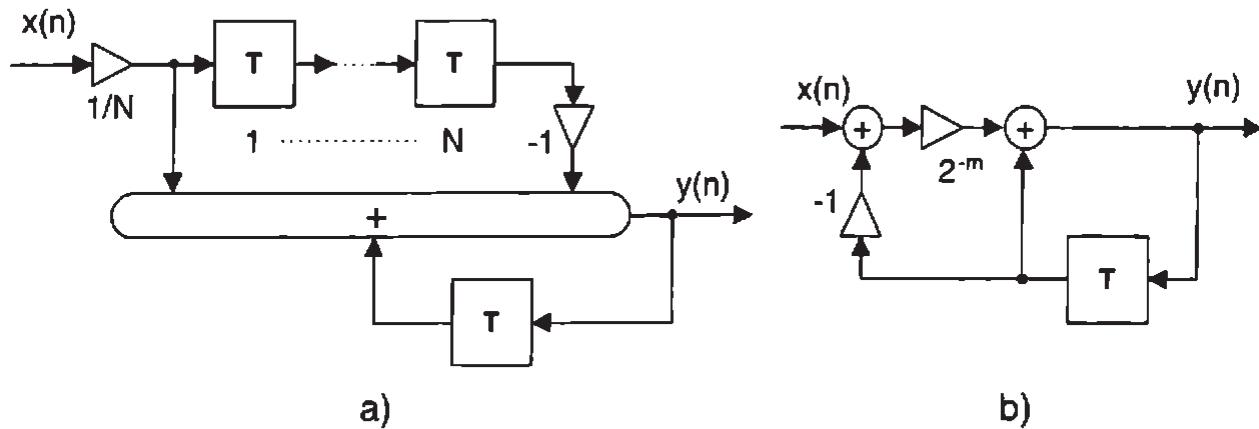
$$\varphi_{MA}(e^{j\theta}) = -(N-1)\frac{\theta}{2}. \quad (4.38)$$



Obrázek 4.14 Frekvenční charakteristika klouzavého průměrovače délky 10

Ze vztahu (4.38) je zřejmé, že jde o filtr s lineární fází. Graf frekvenční charakteristiky je v obr. 4.14.

Z výrazu (4.35) je vidět, že lze filtr MA realizovat také jako rekurzivní filtr. Jde o určitou analogii obvodu z obr. 4.6. Příslušné schéma uvádí obr. 4.15a.



Obrázek 4.15 Nevážený klouzavý průměr jako rekurzivní filtr (a) a vážený klouzavý průměr podle (4.42)

Užitečnou aplikací klouzavého průměru je potlačení rušivého periodického signálu s nulovou stejnosměrnou složkou (typicky sinusového rušení) s frekvencí  $f_r$  superponovaného na konstantní signál. Toho se dosáhne, zvolíme-li délku filtru  $N$  tak, aby první nula frekvenční charakteristiky padla na frekvenci rušení. Musí tedy platit<sup>5</sup>

$$f_{vz} = N f_r . \quad (4.39)$$

Výška postranních laloků amplitudové frekvenční charakteristiky filtru MA klesá úměrně  $1/N$ . Zapojením dvou takových filtrů do kaskády docílíme poklesu úměrnému  $1/N^2$ .

*Klouzavý průměr s exponenciálním zapomínáním* (exponenciálně vážený klouzavý průměr) počítá průměr z  $N$  posledních hodnot, jejichž váhy  $V_n$  exponenciálně klesají se vzdáleností od posledního vzorku (nejstarší vzorek má nejmenší váhu). Jedna z možných variant je  $V_n =$

<sup>5</sup>Tato podmínka je snadněji odvoditelná z průběhu v časové oblasti: délka filtrace  $NT$  musí právě pokrýt období rušivého signálu  $T_r$ .

$V_1^n$ ,  $n \in \{0, N\}$ . Diferenční rovnice filtru je potom

$$y(n) = \frac{1}{\sum_{i=n-(N-1)}^n V_1^{n-i}} \sum_{i=n-(N-1)}^n V_1^{n-i} \cdot x(i). \quad (4.40)$$

Pro  $V_1^2 \ll 1$  platí

$$y(n) \doteq (1 - V_1) \sum_{i=n-(N-1)}^n V_1^{n-i} \cdot x(i).$$

Odpovídající rekurentní vztah je

$$\frac{y(n)}{1 - V_1} - \frac{V_1}{1 - V_1} y(n-1) = x(n) - V_1^N x(n-N),$$

a tedy

$$H(z) = \frac{Y(z)}{X(z)} \doteq \frac{(1 - V_1)(1 - V_1^N z^{-N})}{1 - V_1 z^{-1}}, \quad (4.41)$$

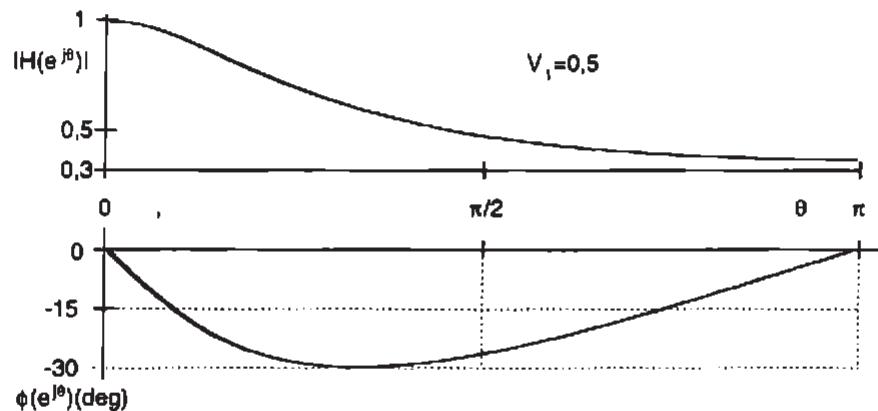
neboť  $V_1^N \ll 1$ . Pro  $V_1 = 0,5$  je tedy

$$H(z) = \frac{1}{2} \frac{1}{1 - 0,5z^{-1}} \quad (4.42)$$

a frekvenční charakteristika tohoto filtru je

$$H(e^{j\theta}) \doteq \frac{1 - V_1}{1 - V_1 e^{-j\theta}}, \quad (4.43)$$

kde  $V_1$  je váha prvního členu, kterou je násoben v každém kroku filtrace každý člen posloupnosti. Průběh frekvenční charakteristiky tohoto filtru uvádí obr. 4.16. Tento typ filtru (s



Obrázek 4.16 Frekvenční charakteristika klouzavého průměru s exponenciálním zapomínáním (pro  $V_1 = 0,5$ )

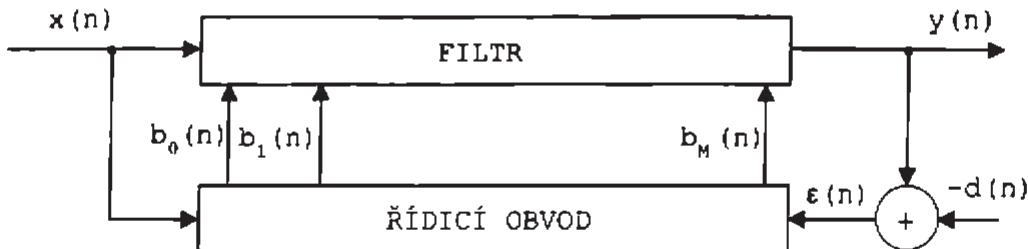
$V_1 = 0,5$ ) se jednoduše realizuje hardwarem při práci s čísly ve dvojkové soustavě (obr. 4.15 b). Tam totiž dělení dvěma znamená pouze posuv čísla o jeden bit vpravo.

#### 4.1.4 Princip adaptivních číslicových filtrů a příklady jejich aplikací

Adaptivní filtry mění své parametry podle určitého kritéria v průběhu filtrace. Používají se v situacích, kdy není předem jasné, jaké by měly být hodnoty koeficientů, aby byla nejlépe vyřešena daná úloha. Navíc se v těchto situacích optimální hodnoty koeficientů mění s časem.

Adaptivní filtr (obr. 4.17) tvoří dvě hlavní části

- vlastní filtr, který je vždy typu FIR, takže sám nemůže být nestabilní,
- řídící obvod, který podle zvoleného kritéria iterativně dostavuje v každém kroku hodnoty koeficientů filtru  $b_i$ .



Obrázek 4.17 Princip adaptivního číslicového filtru

V obr. 4.17 je  $-d(n)$  pomocný řídící (požadovaný) signál. Algoritmus filtru má minimalizovat chybový signál

$$\epsilon(n) = y(n) - d(n) \quad (4.44)$$

Nejznámější z algoritmů je algoritmus minima čtverců odchylek, jehož princip spočívá v iterativním nastavení koeficientů filtru tak, aby bylo dosaženo minima hodnoty výkonu chybového signálu  $\epsilon(n)$ . Jde o minimum hodnoty  $E$ :

$$E = \sum_{n=0}^{N-1} [y(n) - d(n)]^2 = \sum_{n=0}^{N-1} \left[ \sum_{k=0}^M h(k)x(n-k) - d(n) \right]^2. \quad (4.45)$$

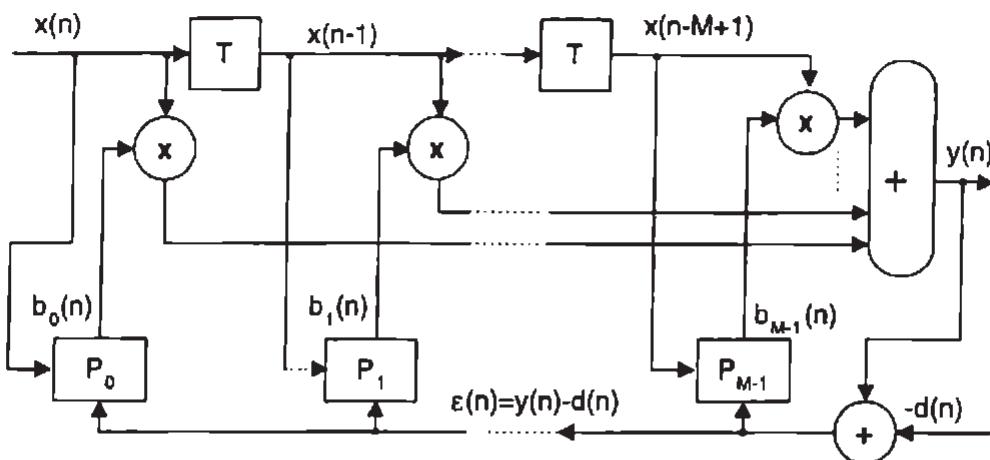
Toto kritérium lze použít k odvození iterativního výpočtu koeficientů. Jedním z možných algoritmů je algoritmus *LMS* (angl. least-mean square algorithm), kde se koeficienty filtru počítají podle vztahů

$$b_i(n+1) = b_i(n) + \Delta b_i(n), \quad (4.46)$$

$$\Delta b_i(n) = \alpha \epsilon(n) x(n-i), \quad i = 0, 1, \dots, N-1. \quad (4.47)$$

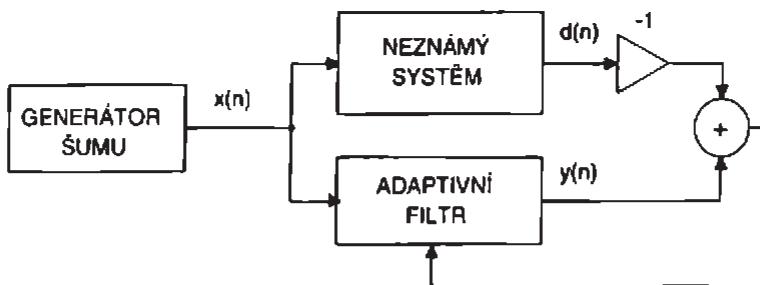
Poté co koeficient dosáhne optimální hodnoty, začne kolem této hodnoty kolísat s amplitudou závislou na algoritmu a konstantách iterativních vztahů pro nastavování koeficientů. Při jejich nevhodné volbě může být adaptivní filtr nestabilní, i když vychází z tranverzálního FIR filtru. Rychlosť konvergence a velikost ustálených fluktuací určuje při použití vztahů (4.46) a (4.47) velikost konstanty  $\alpha$ . Jako první sada koeficientů se může vzít např. nulová hodnota všech koeficientů.

Podrobněji můžeme strukturu adaptivního filtru rozkreslit podle obr. 4.18. Aplikačních možností adaptivních filtrů je velké množství. Je možno je využít pro potlačování „duchů“ v přijímaném televizním signálu (kde se jako signál  $-d(n)$  použije signál zpětného běhu paprsku, jehož nezkreslený průběh je znám), pro kompenzaci zkreslení signálu ve sdělovacím kanálu, pro potlačování šumu (např. rušivého pozadí při komunikaci mobilním telefonem v autě nebo potlačování úzkopásmového rušení s proměnnou frekvencí superponovaného k širokopásmovému signálu), v medicíně k indikaci srdečního tepu plodu rušeného tepem matky, viz např. [15].

Obrázek 4.18 Transverzální adaptivní číslicový filtr. Bloky  $P_i$  realizují (4.46)

Teorie adaptivní filtrace je dnes samostatnou disciplínou zpracování signálů. Podrobnější informace by přesahly rozsah tohoto skripta. Řadu informací lze nalézt v [91]. Zmíníme se pouze o použití adaptivní filtrace pro modelování a identifikaci systémů.

Chceme najít koeficienty filtru, který z daného vstupního signálu vyrobí požadovaný výstupní signál  $y_d(n)$ . Princip znázorňuje obr. 4.19. Pustíme-li signál  $x(n)$  na vstup zkoumaného obvodu z obr. 4.19 a výstupní signál tohoto obvodu použijeme jako žádaný signál  $-d(n)$ , nastaví se postupně koeficienty adaptivního filtru na hodnoty, pro které je rozdíl  $y(n) - d(n)$  minimalizován.



Obrázek 4.19 Modelování systému pomocí adaptivního filtru (podle [39])

## 4.2 Metody návrhu číslicových filtrů

Návrh číslicových filtrů vychází z požadovaného průběhu amplitudové frekvenční charakteristiky filtru  $|H(e^{j\theta})|$ , případně  $|H(f)|$  resp.  $|H(\omega)|$ . Průběh fázové charakteristiky není zadán, nebo se předpokládá lineární fáze. Požaduje-li se lineární fáze, volí se FIR filtr.

Dnes se ve většině případů navrhují číslicové filtry pomocí počítačových programů, bud specializovaných na návrh filtrů, nebo obecnějších, jako například MATLAB, Signal Processing Toolbox. Pokud je takový program k dispozici, může návrh filtru představovat pouze jednořádkový příkaz. Obecně ale je *postup návrhu číslicového filtru* následující:

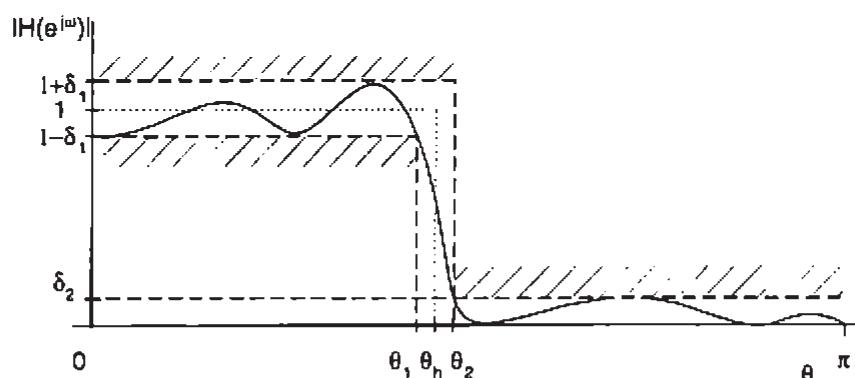
- volba typu filtru - FIR nebo IIR,

- volba řádu filtru (ten v některých případech umožňuje určit procedura programu, např. pro zvolený druh IIR filtru a předepsané tolerance v propustném a závěrném pásmu filtru),
- volba struktury filtru (jde-li o IIR filtr) a výpočet koeficientů přenosu,
- kontrola splnění zadání (výpočet a zobrazení frekvenční charakteristiky filtru).

Pokud frekvenční charakteristika navrženého filtru nevyhovuje předepsaným tolerancím, návrh se opakuje (např. pro vyšší řád filtru).

Programy pro návrh filtrů se liší možnostmi (a cenou). Prakticky všechny umožňují navrhovat základní frekvenčně selektivní filtry (DP, HP, PP, PZ, viz str. 49). Při návrhu se zadává buď typ filtru (FIR nebo IIR), řád filtru a druh filtru (např. pásmová zádrž řádu 5, Čebyševův filtr prvního druhu), nebo je zadán toleranční diagram a druh filtru (např. FIR1 nebo Butterworth) a program sám určí řád filtru a jeho koeficienty tak, aby bylo splněno předepsané toleranční schéma.

Základním typem frekvenčně selektivního filtru je dolní (dolnofrekvenční) propust. Z ní se pomocí tzv. frekvenčních transformací počítají koeficienty ostatních typů frekvenčně selektivních filtrů. Způsob těchto transformací je odlišný pro FIR filtry a IIR filtry a bude stručně popsán dále. Příklad tolerančního diagramu dolní propusti uvádí obr. 4.20. V něm je rozsah frekvencí  $\langle 0, \theta_1 \rangle$  propustné pásmo filtru, rozsah frekvencí  $\langle \theta_1, \theta_2 \rangle$  přechodné pásmo filtru a rozsah frekvencí  $\langle \theta_2, \pi \rangle$  nepropustné pásmo filtru.



Obrázek 4.20 Toleranční diagram dolní propusti

V praxi jsou koeficienty přenosu filtrů reálná čísla (impulsní odezva je reálná). Proto je amplitudová frekvenční charakteristika sudá funkce  $\theta$  a fázová frekvenční charakteristika lichá funkce  $\theta$ . Tyto charakteristiky jsou navíc periodické s periodou  $2\pi$ . Proto stačí zadat jejich průběh pouze v polovině základního intervalu  $\langle 0, \pi \rangle$  (v případě horizontální osy popsané v hodnotách  $f$  v intervalu  $(0, f_{vz}/2)$ ). Uvedené skutečnosti demonstrují obr. 4.12 a 4.13. Je-li při vzorkování signálu dodržena vzorkovací věta (v případě frekvenčně neomezeného signálu zapojením antialiasingové analogové dolní propusti na vstup systému), je signál pouze v pásmu  $\theta$  od 0 do  $\pi$  a periodičnost frekvenční charakteristiky nevadí.

Protože jsou v současné době číslicové filtry navrhovány téměř výlučně počítačem, budou zde probrány pouze základní metody návrhu. Domníváme se totiž, že by uživatel měl mít představu o tom, jak koeficienty filtru počítá použitý program.

### 4.2.1 Návrh filtrů s konečnou impulsní odezvou

Zde má základní význam průběh impulsní odezvy. Její hodnoty jsou přímo koeficienty transverzálního FIR filtru (obr. 4.2). Z tvaru impulsní odezvy je také přímo vidět, zda jde o filtr s lineární fází (obr. 4.4).

Protože u FIR filtru prakticky vždy požadujeme lineární fází, navrhují se tyto filtry tak, že jejich impulsní odezvu volíme jako sudou nebo lichou posloupnost  $n$ . Filtr s takovou odezvou je nekauzální a má fázi 0 nebo  $\pm\pi/2$ . Chceme-li filtrovat signál v reálném čase (dostat vzorek výstupu před odebráním dalšího vzorku vstupu), musíme posunout impulsní odezvu o polovinu její délky (o  $(M+1)/2$ ) doprava. Tomu odpovídá ve frekvenční oblasti násobení komplexní exponenciálou (viz (3.27)), takže amplitudová frekvenční charakteristika se nezmění, ale fázová frekvenční charakteristika přejde v přímluku se zápornou směrnicí.

#### Metoda Fourierovy transformace diskrétních signálů (FTD) a oken

Je základní metodou návrhu FIR filtrů<sup>6</sup>. Při návrhu se vyjde z požadované frekvenční charakteristiky filtru  $H_d(e^{j\theta})$ . Pomocí IFTD najdeme požadovaný průběh impulsní odezvy  $h_d(n)$ . Tato odezva je nekauzální a nekonečně dlouhá<sup>7</sup>.

Protože navrhujeme FIR filtr, vezmeme z  $H_d(n)$  konečný počet členů  $L = M + 1$ , kde  $M$  je požadovaný řád filtru. To odpovídá násobení  $h_d(n)$  obdélníkovým oknem  $w_o(n)$  délky  $L$ . Frekvenční charakteristiku filtru pak dostaneme složením konečného počtu členů Fourierovy řady a v místech skokových změn ideální (požadované) frekvenční charakteristiky se projeví překývnutí, tzv. *Gibbsův jev*. Toto překývnutí nezmizí ani při růstu  $L$  - zmenšuje se pouze šířka základního okna, ale výška maximálního výkyvu zůstává přibližně konstantní. Ve frekvenční oblasti vyvolává toto zakmitání konvoluce požadované frekvenční charakteristiky se spektrem obdélníkového okna.

Postup je tedy následující. Je dána požadovaná frekvenční charakteristika  $H_d(e^{j\theta})$ , pro niž platí

$$H_d(e^{j\theta}) = \sum_{n=-\infty}^{\infty} h_d(n)e^{-jn\theta}. \quad (4.48)$$

Odtud dostaneme požadovanou impulsní odezvu pomocí IFTD:

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\theta}) e^{jn\theta} d\theta. \quad (4.49)$$

Posloupnost  $h_d(n)$  vynásobíme obdélníkovým oknem

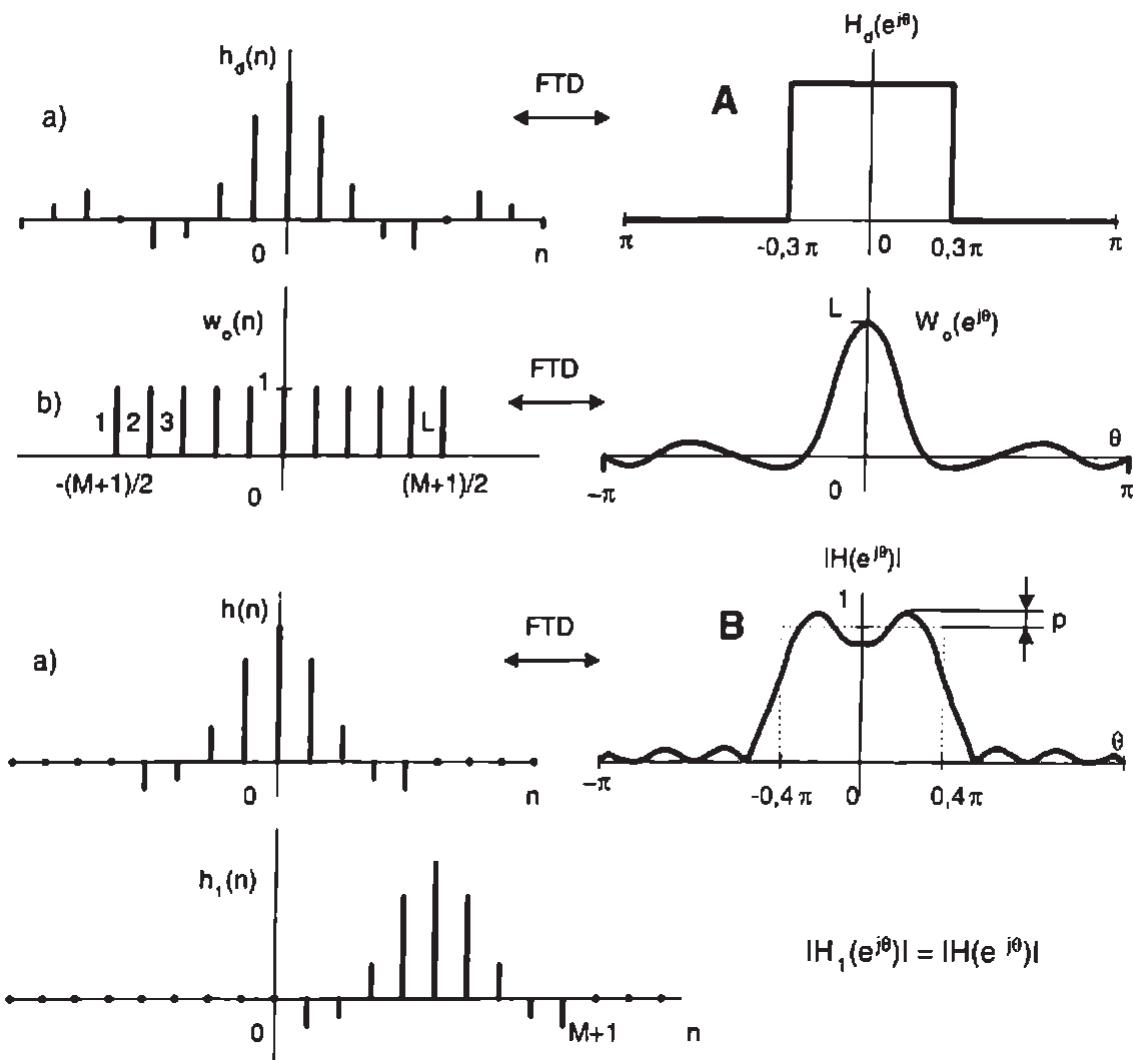
$$w_o(n) = \begin{cases} 1 & \text{pro } -M/2 \leq n \leq M/2, \\ 0 & \text{jinde.} \end{cases} \quad (4.50)$$

Impulsní odezva navrženého FIR filtru je

$$h(n) = h_d(n) \cdot w_o(n). \quad (4.51)$$

Ve frekvenční oblasti odpovídá násobení dle (4.51) konvoluce

$$H(e^{j\theta}) = H_d(e^{j\theta}) * W_0(e^{j\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\nu}) \cdot W_0(e^{j(\theta-\nu)}) d\nu. \quad (4.52)$$



Obrázek 4.21 Návrh dolní propusti FIR metodou FTD a oken. Výchozí charakteristika označena A, výsledná B

Graficky vyjadřuje postup návrhu obr. 4.21.

Zvlnění výsledné charakteristiky  $B$  z obr. 4.21 je výsledek konvoluce výchozí charakteristiky se spektrem obdélníkového okna. Překmit amplitudy  $p$  zůstává zhruba konstantní i při růstu rádu filtru  $M$ . Potlačení zákmitů na skocích frekvenční charakteristiky se dosáhne použitím jiných oken pro násobení požadované impulsní odezvy. Volí se okna s nižšími postrannými laloky než má okno obdélníkové. Hodnota překmitu  $p$  ve výsledné charakteristice v obr. 4.21 se podstatně změní nebo prakticky zmizí. Cenou za to je ale rozšíření přechodného pásma filtru, odpovídající větší šířce hlavního laloku použitého okna proti oknu obdélníkovému. *Obdélníkové okno* je definováno vztahem (4.53).

Dále uvedeme *definice oken* nejpoužívanějších při návrhu číslicových filtrov. Tato okna se ovšem využívají i v FFT spektrální analýze. Okna budou uvedena v *kauzální formě* (tedy pro

<sup>6</sup>V MATLAB Signal Processing Toolbox je tato metoda užita při návrhu filtrov příkazy *firl*, *fir2*, *firls*.

<sup>7</sup>Jde totiž o Fourierovu řadu periodické frekvenční charakteristiky filtrov. Do Fourierovy řady se zde rozvíjí funkce frekvence, nikoliv času. Je-li frekvenční charakteristika reálná a sudá, je i impulsní odezva reálná a sudá.

nezáporná  $n$ ). Pro kauzální obdélníkové okno platí

$$w_0(n) = \begin{cases} 1 & \text{pro } 0 \leq n \leq M, \\ 0 & \text{jinde.} \end{cases} \quad (4.53)$$

- Okno trojúhelníkové (Bartlett):<sup>8</sup>

$$w_T(n) = \begin{cases} \frac{n}{M/2}, & 0 \leq n \leq M/2, \\ 2 - \frac{2n}{M}, & M/2 \leq n \leq M, \\ 0 & \text{jinde.} \end{cases} \quad (4.54)$$

- Okno Hann (hanning, Hanning, von Hann):

$$w_{HN}(n) = \begin{cases} \frac{1}{2}(1 - \cos \frac{2\pi n}{M}), & 0 \leq n \leq M, \\ 0 & \text{jinde.} \end{cases} \quad (4.55)$$

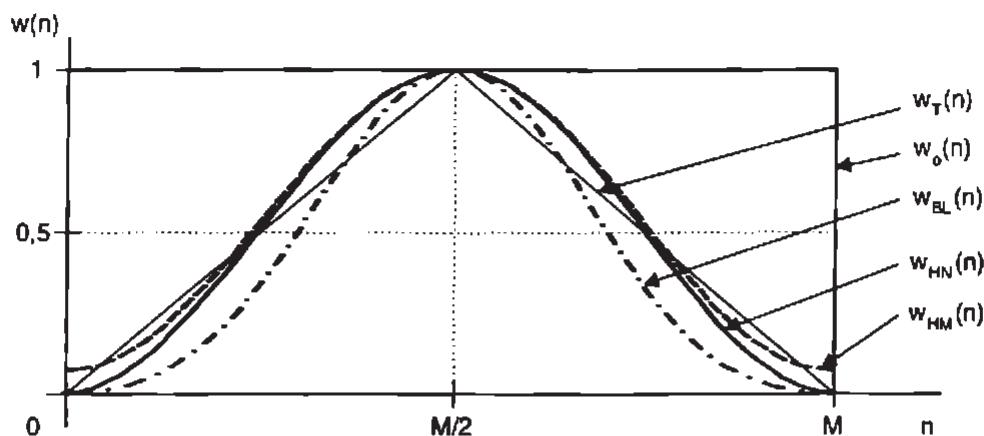
- Okno Hamming:

$$w_{HM}(n) = \begin{cases} 0,54 - 0,46 \cdot \cos \frac{2\pi n}{M}, & 0 \leq n \leq M, \\ 0 & \text{jinde.} \end{cases} \quad (4.56)$$

- Okno Blackman:

$$w_{BL}(n) = \begin{cases} 0,42 - 0,50 \cdot \cos \frac{2\pi n}{M} + 0,08 \cdot \cos \frac{4\pi n}{M}, & 0 \leq n \leq M, \\ 0 & \text{jinde.} \end{cases} \quad (4.57)$$

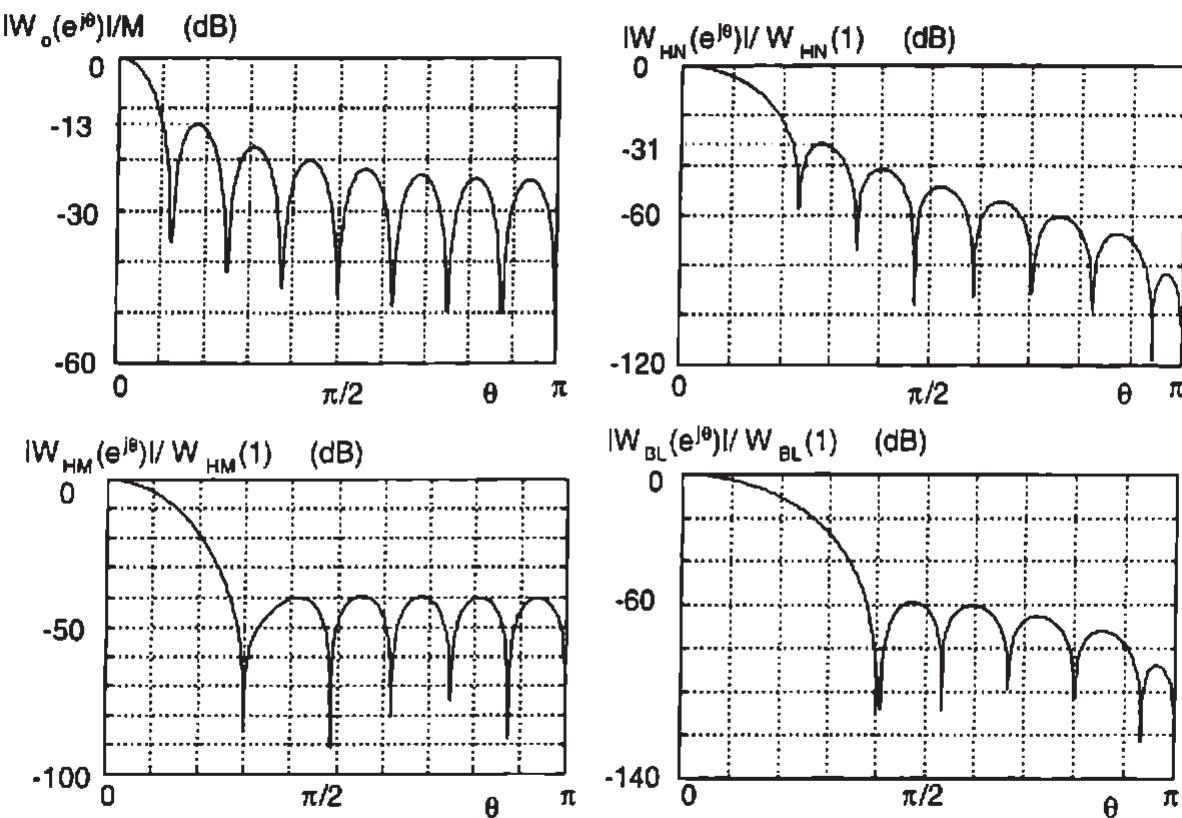
Průběhy těchto oken v časové oblasti (obálky čárového průběhu) uvádí obr. 4.22, průběhy spekter těchto oken uvádí obr. 4.23. Frekvenční spektra oken se obvykle kreslí v semilogarit-



Obrázek 4.22 Průběhy základních oken v časové oblasti

mických souřadnicích. Počet jejich vln (laloků) závisí na délce okna  $L = M + 1$ , ale vzájemný poměr výšek sousedních vln (tvar spektra) závisí na  $M$  málo. Šířka hlavního laloku je nejmenší u obdélníkového okna, u ostatních výše uvedených oken je dvojnásobná nebo trojnásobná.

<sup>8</sup>V MATLAB Signal Processing Toolbox se definice okna Bartlettova od trojúhelníkového liší tím, že trojúhelníkové má nenulovou první a poslední hodnotu.



Obrázek 4.23 Normovaná spektra základních oken (obdélníkového, Hannova, Hammingova a Blackmanova) pro  $M + 1 = 16$

Tabulka 4.1 Klasická okna: základní vlastnosti a užití pro návrh FIR DP

název okna	1. postranní lalok	šířka hl. laloku	šířka přech. pásma	minimální zeslabení
obdélník	-13dB	$4\pi/(M+1)$	$1.8\pi/(M+1)$	21 dB
trojúhelník	-25dB	$8\pi/(M+1)$	$5.6\pi/(M+1)$	25 dB
Hann	-31dB	$8\pi/(M+1)$	$6.2\pi/(M+1)$	44 dB
Hamming	-41dB	$8\pi/(M+1)$	$6.6\pi/(M+1)$	53 dB
Blackman	-57dB	$12\pi/(M+1)$	$11\pi/(M+1)$	74 dB

Použijeme-li výše uvedená okna pro návrh filtrů, je možno zvolit typ okna pro požadované minimální zeslabení v nepropustném pásmu a délku filtru  $M + 1$  pro předepsanou šířku přechodného pásma podle tabulky 4.2.1 (dle [39]).

Uvedenou metodou lze navrhovat různé typy FIR filtrů. Ukážeme návrh dolní propusti a derivátoru.

*Návrh dolní propusti s mezní frekvencí  $\theta_h$  metodou FTD a oken*

Plati

$$H_d(e^{j\theta}) = \begin{cases} 1, & |\theta| \leq \theta_h, \\ 0, & |\theta| < \pi. \end{cases} \quad (4.58)$$

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\theta}) e^{jn\theta} d\theta = \frac{1}{2\pi} \int_{-\theta_h}^{\theta_h} e^{jn\theta} d\theta = \frac{\theta_h}{\pi} \frac{\sin(n\theta_h)}{n\theta_h}. \quad (4.59)$$

Frekvenční charakteristiku DP zadáváme jako reálnou (a sudou). Takový filtr má reálnou a sudou impulsní odezvu a je tedy nekauzální. Chceme-li kauzální filtr, pak po nalezení  $h(n)$  posuneme tuto funkci o  $n_0 = M/2$  (resp.  $n_0 = (M+1)/2$  pro  $M$  liché) doprava. Tomuto posuvu odpovídá násobení  $H(e^{j\theta})$  exponenciálou  $e^{-j(M/2)\theta}$ , takže dostáváme kauzální filtr s lineární fází. Přitom  $(M/2)T$  odpovídá skupinovému zpoždění, čili signál v propustném pásmu filtru je zpožděn o  $M/2$  vzorkovacích kroků.

Doporučený rozsah hodnot  $\theta_h$  je  $0,3\pi$  až  $0,5\pi$ , čili  $f_{vz}/f_h \in (4,6)$ .

Obdobným postupem je možné odvodit přímo výrazy pro frekvenční charakteristiky FIR filtrů typu HP, případně PP a PZ s mezními frekvencemi  $\theta_d$  a  $\theta_h$ . Porovnáním výrazu pro frekvenční charakteristiku DP s výrazy pro HP, PP a PZ lze navíc odvodit transformační rovnice pro převod koeficientů DP na koeficienty HP, PP a PZ (viz např. [11]). My ale uvedeme dále jinou metodu této transformace.

**Příklad 4.1** Určete koeficienty dolní propusti řádu  $M = 6$  s mezní frekvencí  $f_h = 250$  Hz při vzorkovací frekvenci  $f_{vz} = 1$  kHz tak, aby šlo o filtr s lineární fází. ▲

**Řešení**

$$\theta_h = 2\pi \frac{f_h}{f_{vz}} = \frac{\pi}{2}.$$

Koeficienty filtru získáme z (4.59), přičemž vyžadujeme sudou impulsní odezvu. Najde me tedy

$$h(n) = b_n = h(-n) = b_{-n} = 0,5 \cdot \frac{\sin(n\pi/2)}{n\pi/2}, \quad |n| \leq 3.$$

Odtud

$$\begin{aligned} h(0) &= b_0 = 0,5; \quad h(1) = b_1 = b_{-1} = 1/\pi; , \\ h(2) &= b_2 = b_{-2} = 0; \quad h(3) = b_3 = b_{-3} = -1/(3\pi) . \end{aligned}$$

Je tedy

$$h(n) = -\frac{1}{3\pi} \delta(n-3) + \frac{1}{\pi} \delta(n-1) + 0,5 \delta(n) + \frac{1}{\pi} \delta(n+1) - \frac{1}{3\pi} \delta(n+3).$$

Odtud

$$H(z) = 0,5 + \frac{1}{\pi}(z + z^{-1}) - \frac{1}{3\pi}(z^3 + z^{-3}).$$

Dosazením  $z = e^{j\theta}$  dostáváme výraz pro frekvenční charakteristiku:

$$H(e^{j\theta}) = 0,5 + \frac{2}{\pi} \cos \theta - \frac{2}{3\pi} \cos(3\theta).$$

**Příklad 4.2** Určete metodou FTD a okén koeficienty číslicového derivátoru. ▲

**Řešení** Ideální derivátor má lichou a imaginární frekvenční charakteristiku (4.26). Jeho impulsní odezva je tedy lichá a reálná (viz (3.74)). Platí

$$\begin{aligned} h_d(n) &= b_n = -h_d(-n) = -b_{-n}, \quad n \leq (M/2), \\ h_d(n) &= b_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} j\theta e^{jn\theta} d\theta. \end{aligned} \quad (4.60)$$

Pro  $n = 0$  je  $h_d(0) = 0$ . Integrací per partes nebo užitím tabulek integrálů dostaneme vztah pro  $\int xe^{ax} dx$  a odtud dostaneme

$$b_n = \frac{j}{2\pi} e^{jn\theta} \left[ \frac{\theta}{jn} - \frac{1}{(jn)^2} \right] \Big|_{-\pi}^{\pi} = \frac{(-1)^n}{n}. \quad (4.61)$$

Např. pro  $M = 6$  dostaneme diferenční rovnici

$$y(n) = \frac{1}{3}x(n+3) - \frac{1}{2}x(n+2) + x(n+1) - x(n-1) + \frac{1}{2}x(n-2) - \frac{1}{3}x(n-3).$$

Té odpovídá přenos

$$H(z) = \frac{1}{3}(z^3 - z^{-3}) - \frac{1}{2}(z^2 - z^{-2} + (z - z^{-1})),$$

a odpovídající frekvenční odezva je

$$H(e^{j\theta}) = 2j \left[ \frac{1}{3} \sin(3\theta) - \frac{1}{2} \sin(2\theta) + \sin\theta \right]. \quad (4.62)$$

Tato charakteristika vykazuje značné zvlnění, které lze odstranit vynásobením výchozí impulsní odehyvy např. oknem Hann.

Pro  $M = 20$  dostaneme koeficienty filtru

$$h_D(n) = b_n = \frac{(-1)^n}{n}, \quad 0 < n \leq 10. \quad (4.63)$$

S úpravou oknem Hann dostaneme koeficienty ze vztahu

$$h_{DW} = b_n = \frac{(-1)^n}{n} \cdot 0,5 \cdot \left( 1 - \cos \frac{2\pi n}{10} \right), \quad 0 < n \leq 10. \quad (4.64)$$

S výjimkou trojúhelníkového všechna dosud uvedená okna patří do třídy tzv. *zobecněných kosinových oken*, tvořených součtem několika kosinusovek s frekvencemi  $k \cdot 2\pi/M$ . Jsou v praxi nejužívanější a jejich spektra jsou tvořena lieárními kombinacemi posunutých spekter obdélníkového okna. Kromě nich existují složitější okna, např. okno Dolphovo-Čebyševovo a okno Kaiserovo.

*Kaiserovo okno* je zajímavé tím, že vhodnou volbou svých dvou parametrů umožňuje velmi dobře approximovat ostatní klasická okna. Je definováno vztahem

$$w_K(n) = \begin{cases} \frac{I_0[\beta(1 - ((n - \alpha)/\alpha)^2)^{1/2}]}{I_0(\beta)} & \text{pro } 0 \leq n \leq M, \\ 0 & \text{jinde,} \end{cases} \quad (4.65)$$

kde  $\alpha = M/2$  a  $I_0(.)$  je modifikovaná Besselova funkce prvního druhu. Okno má tedy dva parametry - délku  $(M + 1)$  a parametr tvaru  $\beta$ . Lze je využívat jak při návrhu filtrů, tak v FFT spektrální analýze. Více o něm viz [68].

### Nalezení koeficientů základních frekvenčně selektivních filtrů z koeficientů dolní propusti

Tyto koeficienty můžeme najít pomocí metody publikované v [85]. Využijeme linearity FTD. Platí

$$h_1(n) \pm h_2(n) \longleftrightarrow H_1(e^{j\theta}) \pm H_2(e^{j\theta}).$$

Pokud tedy můžeme získat frekvenční charakteristiku požadovaného filtru jako rozdíl nebo součet frekvenčních charakteristik známých filtrů, dostaneme impulsní odezvu (a tedy i koeficienty transverzálního FIR filtru) hledaného filtru jako rozdíl nebo součet impulsních charakteristik oněch známých filtrů.

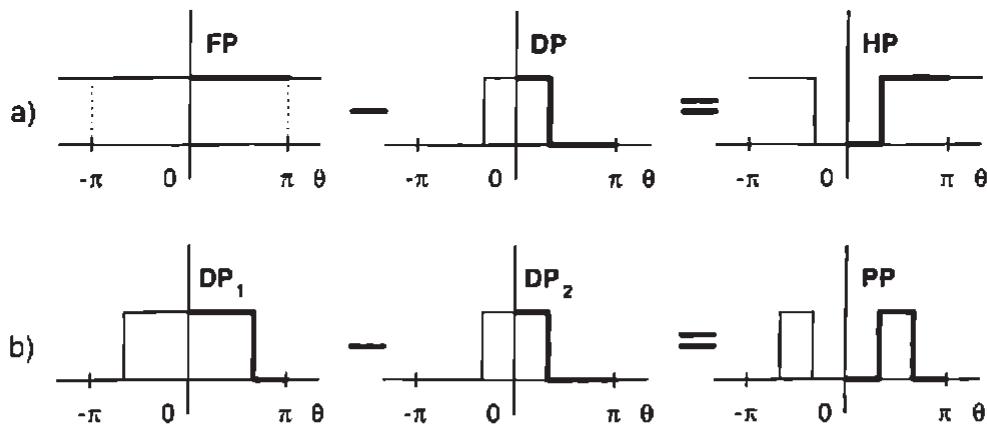
Jako jeden ze známých filtrů vezmeme speciální „fázový posouvač“ s nulovým fázovým posuvem, pro který je výstižnější anglický název „all-pass“ a pro který platí

$$H_{FP}(e^{j\theta}) = |H_{FP}(e^{j\theta})| = 1, \quad h_{FP}(n) = \delta(n), \quad (4.66)$$

čili jeho impulsní odezva je tvořena jediným jednotkovým impulsem pro čas  $n = 0$ . Ta tvoří současně koeficienty filtru, takže platí

$$b_{i,FP} = \begin{cases} 1, & i = 0, \\ 0, & i \neq 0. \end{cases} \quad (4.67)$$

Způsob, jakým z koeficientů DP získáme koeficienty HP a jak z koeficientů dvojice DP dostaneme koeficienty PP, ukazuje obr. 4.24. Koeficienty PZ získáme odečtením koeficientů PP od koeficientů FP.



Obrázek 4.24 Získání frekvenčních charakteristik horní propusti HP (a) a pásmové propusti PP(b) jako rozdíl frekvenčních charakteristik vhodných filtrů

### Návrh FIR filtrů pomocí počítačových programů

Současné počítačové programy pro návrh číslicových FIR filtrů využívají i dokonalejší algoritmy než metodu FTD a oken. Metoda FTD a oken poskytuje při obdélníkovém oknu nejlepší aproximaci požadované frekvenční charakteristiky ve smyslu minima středně kvadratické odchylky požadované a skutečné charakteristiky. Největší odchylky požadované a skutečné charakteristiky jsou přitom v okolí skokových změn požadované charakteristiky.

Jednou z dalších možností návrhu FIR filtrů pomocí počítače je jednoduchá *metoda frekvenčního vzorkování* [15, 11]. Vybereme  $N$  ekvidistantních hodnot (vzorků) požadované frekvenční odezvy a na takto získanou posloupnost aplikujeme IDFT. Nalezené vzorky  $h(n)$  jsou

komplexní. Vezmeme jejich absolutní hodnoty a doplníme průběh  $h(n)$  na jeden ze čtyř typů symetrie pro filtry s lineární fází (obr. 4.4). U této metody sice výsledná frekvenční charakteristika souhlasí s požadovanou ve zvolených bodech, mezi nimi ale může značně kolísat. Velikost oscilací lze snížit vynásobením impulsní odezvy vhodným oknem.

Zřejmě nejpoužívanější metodou návrhu FIR filtrů je dnes algoritmus Parkse a McClellana (viz [73, 74, 68]). Tato metoda využívá Čebyševovy approximace, minimalizuje největší chybu (tzv. „minmax method“) a zaručuje konstantní zvlnění výsledné charakteristiky v propustném a nepropustném pásmu. Frekvenční charakteristika je approximována polynomem v  $\cos \theta$ . Algoritmus je podrobně popsán v [68] a pro zadané toleranční pásmo dává nejnižší řád filtru a konstantní zvlnění v nepropustném pásmu. Posouzení efektivnosti tohoto algoritmu umožňuje následující srovnání. Mějme navrhnutou pásmovou propust pro vzorkovací frekvenci  $f_{uz} = 4\text{ kHz}$ , závěrné pásmo  $0 \div 0,6\text{ kHz}$  a  $2,4 \div 4\text{ kHz}$ , propustné pásmo  $1 \div 2\text{ kHz}$  a zeslabení v závěrném pásmu minimálně  $40\text{ dB}$ . Při použití metody FTD a oken (použito okno Hann) vyjde řád filtru  $M = 63$ , zatímco metodou Parks-McClellan vyjde  $M = 42$ .

#### 4.2.2 Návrh filtrů s nekonečnou impulsní odezvou

Základní výhodou IIR filtrů proti FIR filtrům je to, že zadané specifikace obvykle splní při daleko nižším rádu filtru. To je způsobeno approximováním požadované charakteristiky racionalní lomenou funkcí. Proto méně zpožďuje signál při filtraci. Jejich fázová charakteristika je nelineární. Při zpracování signálu „off-line“ (když jsou všechna data pro filtraci k dispozici před zpracováním) je možno eliminovat nelineární fázové zkreslení vhodným zpracováním<sup>9</sup>.

Klasickou metodou návrhu, kterou se budeme zabývat v tomto skriptu, je *návrh pomocí analogového vzorového filtru* (prototypu). Vychází se z tabulek koeficientů analogového filtru a pomocí konformního zobrazení roviny  $p$  (ve které se pomocí Laplaceovy transformace vyjadřuje přenos analogového filtru) do roviny  $z$  se najdou koeficienty odpovídajícího číslicového filtru. Použitá transformační rovnice musí zobrazovat rovinu  $p$  do roviny  $z$  tak, aby levá polovina  $p$  odpovídala vnitřku jednotkové kružnice v rovině  $z$ , případně jeho části. Pak se při transformaci zachová stabilita filtru. Základní metody pro uvedenou transformaci jsou *metoda invariantnosti impulsní odezvy*, *metoda náhrady derivací diferencemi* a *metoda bilineární transformace*. Poslední metoda je nejrozšířenější, protože kromě zachování stability vylučuje aliasing. Transformace frekvencí z analogové do číslicové oblasti je ale nelineární. To známená, že během návrhu musíme přepočítat mezní frekvence filtru.

IIR filtry se používají zejména jako základní frekvenčně selektivní filtry (obr. 4.12). Návrh většinou vychází z dolní propusti. Její koeficienty se pomocí frekvenčních transformací převedou na koeficienty požadovaného filtru (dolní propusti s jinou mezní frekvencí nebo jiného typu filtru). Tuto transformaci lze provést u analogového filtru a pak teprve přejít k číslicovému filtru, nebo nejprve najít k analogové dolní propusti vzorovou číslicovou dolní propust a transformaci provést až pro filtr číslicový. My se zmíníme dále o této druhé metodě.

IIR filtry je ale možné taky navrhovat *přímou metodou návrhu*, kdy číslicový filtr approximuje požadovanou amplitudovou frekvenční charakteristiku přímkovými úseků. Tak se navrhují filtry s více propustnými nebo závěrnými pásmeny. Této metodě se zde věnovat nebudeme<sup>10</sup>.

<sup>9</sup>V programu MATLAB Signal Processing Toolbox funkci *filtfilt*.

<sup>10</sup>V programu MATLAB Signal Processing Toolbox se tento návrh provádí funkci *yulewalk*.

### Základní analogové dolní propusti

Protože návrh IIR filtrů většinou vychází z analogové dolní propusti, zopakujeme zde stručně základní typy těchto propustí a jejich vlastnosti.

#### Butterworthova dolní propust

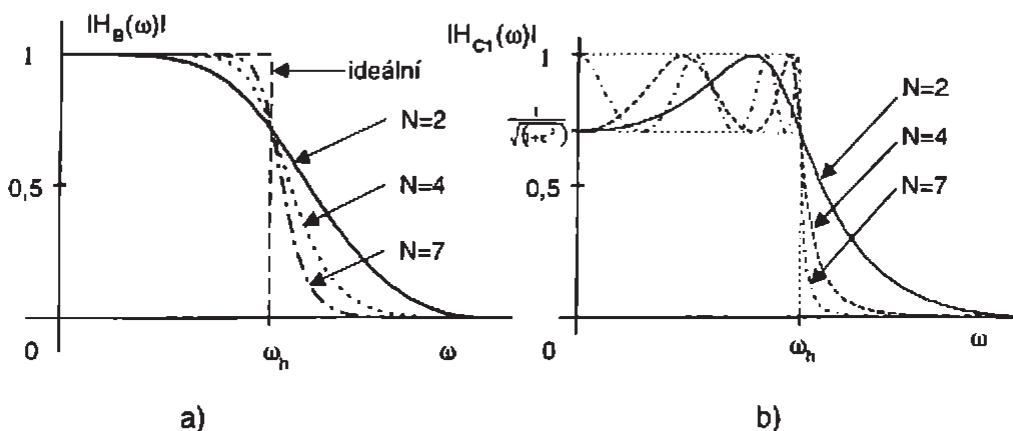
$|H(\omega)|$  je v propustném pásmu maximálně plochá, celý její průběh je monotónní a strmost v přechodném pásmu roste s řádem filtru  $N$  (pokles logaritmické asymptotické frekvenční charakteristiky je v přechodném pásmu  $N \cdot 20 \text{ dB/dek}$ ). Platí

$$|H(0)| = 1; \quad |H(\omega_h)| = 1/\sqrt{2}; \quad |H(\infty)| = 0.$$

Filtr nemá nuly přenosu mimo nekonečno. V časové oblasti odezva tohoto filtru na skok vyzkoušejte poměrně velké zákmity. Pro amplitudovou frekvenční charakteristiku Butterworthovy dolní propusti řádu  $N$  platí

$$|H_A(\omega)|^2 = \frac{1}{1 + (\omega/\omega_h)^{2N}}. \quad (4.68)$$

Frekvenční charakteristiky Butterworthových filtrů tří různých řádů ukazuje obr. 4.25a.



Obrázek 4.25 Frekvenční charakteristiky analogových prototypů IIR filtrů Butterworth (a) a Čebyšev1 (b) vyznačených řádů

#### Čebyševova dolní propust

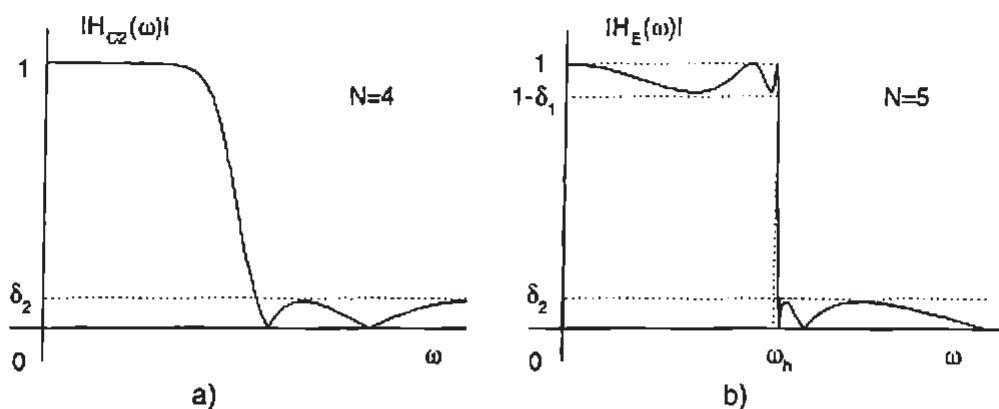
Splnění daného tolerančního diagramu zajistí při nižším řádu než filtr Butterworthův. Amplitudová frekvenční charakteristika  $|H(\omega)|$  má definované rovnoměrné zvlnění buď v propustném pásmu (*Čebyševův filtr prvního typu*), nebo v závěrném pásmu (*Čebyševův filtr druhého typu*). Přechodné pásmo je užší než u filtru Butterworthova. Skoková odezva v časové oblasti má velké překmity. Frekvenční charakteristiky Čebyševových filtrů prvního druhu tří různých řádů ukazuje obr. 4.25b.

Amplitudová frekvenční charakteristika dolní propusti Čebyšev typ 1 je dáná vztahem

$$|H_A(\omega)|^2 = \frac{1}{1 + \epsilon^2 \cdot T_N^2(\omega/\omega_h)}, \quad (4.69)$$

kde  $\epsilon$  určuje zvlnění a  $T_N(\omega)$  je Čebyševův polynom řádu  $N$ . Pro ten platí

$$T_N(\omega) = \cos(N \cdot \cos \omega). \quad (4.70)$$



Obrázek 4.26 Frekvenční charakteristiky analogových prototypů IIR filtrů Čebyšev2 (a) a Cauer

Z obr. 4.25 b je vidět, že pro  $\omega > \omega_h$  je průběh  $|H_A(\omega)|$  monotónní. Platí  $|H_A(0)| = 1$  pro  $N$  liché a  $|H_A(0)| = (1 + \varepsilon^2)^{-1/2}$  pro  $N$  sudé. Také tento filtr má nuly pouze v nekonečnu.

Příklad frekvenční charakteristiky Čebyševova filtru druhého druhu uvádí obr. 4.26 a.

#### Cauerova (eliptická) dolní propust

V r. 1931 ukázal Cauer, že nejlepší approximace ideální dolní propusti má nuly pro konečné  $\omega$ . Při vhodné volbě těchto nul se dosáhne rovnoměrného zvlnění v propustném i nepropustném pásmu. Souřadnice těchto nul se určují pomocí tzv. eliptických funkcí.

$|H(\omega)|$  má definované zvlnění v obou pásmech ( $\delta_1, \delta_2$ ) a pro dané  $N, \omega_h, \delta_1, \delta_2$  má nejužší přechodné pásmo ze všech filtrů. Platí

$$|H_A(\omega)|^2 = \frac{1}{1 + \varepsilon^2 \cdot R_N^2(\omega/\omega_h)}, \quad (4.71)$$

$$R_N(\omega) = \frac{\omega(\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2) \cdots (\omega_k^2 - \omega^2)}{(1 - \omega_1^2 \omega^2)(1 - \omega_2^2 \omega^2) \cdots (1 - \omega_k^2 \omega^2)}, \quad k = \begin{cases} N/2, & N\text{-sudé}, \\ (N-1)/2, & N\text{-liché}. \end{cases} \quad (4.72)$$

Tyto filtry mají nejlepší vlastnosti z hlediska amplitudové frekvenční charakteristiky, ale obtížně se navrhují. K návrhu se používá počítač. Příklad frekvenční charakteristiky Cauerova filtru uvádí obr. 4.26 b.

Dalším typem dolní propusti je *Besselova dolní propust*. Má sice nejmenší překmity ve skokové odezvě v časové oblasti a její fázová charakteristika se nejvíce blíží lineární, ale pro daný řád filtru má nejsírší přechodné pásmo. Bližší informace viz např. [52].

#### Bilineární transformace

Bilineární transformace je nejpoužívanější metoda přechodu od přenosu  $H_A(p)$  vzorového analogového filtru k přenosu číslicového filtru  $H(z)$ . Odvození transformačního vztahu je založeno na integraci diferenciální rovnice popisující analogový filtr. Integrace se přitom provádí pomocí lichoběžníkového pravidla.

Odvození ukážeme na případu ideálního analogového derivátoru. Ten je za nulových počátečních podmínek popsán vztahy

$$y(t) = \frac{dx(t)}{dt}, \quad (4.73)$$

$$H_A(p) = \frac{Y(p)}{X(p)} = p. \quad (4.74)$$

Integrací (4.73) dostaneme

$$x(t) = \int_{t_0}^t y(\tau) d\tau,$$

a pro  $t_0 = (n-1)T$ ,  $t = nT$  a integraci lichoběžníkovým pravidlem (viz (4.23) a následující poznámka) dostaváme

$$x(nT) - x[(n-1)T] = \frac{T}{2} \{y(nT) + y[(n-1)T]\}.$$

Při obvyklém zápisu tedy

$$x(n) - x(n-1) = \frac{T}{2}[y(n) + y(n-1)].$$

Transformací  $Z$  této rovnice dostaneme

$$H(z) = H_A(p)|_{p=\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}}. \quad (4.75)$$

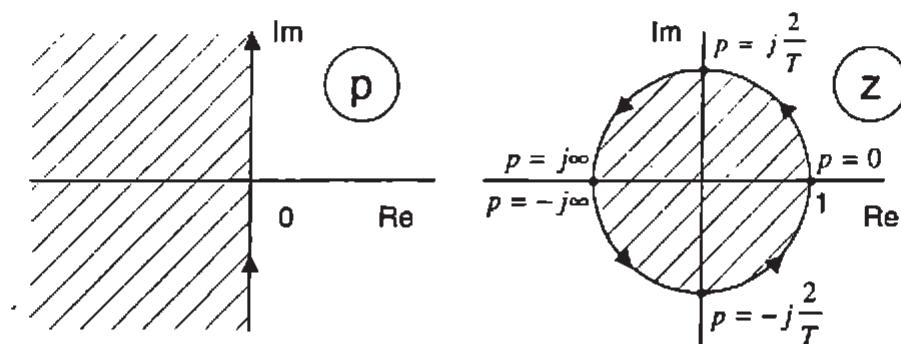
Vztah

$$p = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} \quad (4.76)$$

a inverzní vztah

$$z = \frac{2+pT}{2-pT} \quad (4.77)$$

definuje konformní zobrazení roviny  $p$ <sup>11</sup> do roviny  $z$  takové, že imaginární osa roviny  $p$  odpovídá jednotková kružnice v rovině  $z$  a levé polovině  $p$  odpovídá vnitřek jednotkového kruhu v  $z$  (obr. 4.27). Ze stabilního analogového filtru dostaneme tedy bilineární transformaci stabilní filtr číslicový. Konformnost zobrazení dle (4.77) znamená, že je zobrazení vzájemně jednoznačné a zachovává úhly.



Obrázek 4.27 Zobrazení roviny  $p$  do roviny  $z$  pomocí bilineární transformace

Z obr. 4.27 je zřejmé, že celý rozsah analogových frekvencí  $-\infty < \omega_A < \infty$  se zobrazí do základního intervalu frekvencí číslicového filtru  $-\pi < \theta \leq \pi$ , protože  $\theta$  odpovídá úhlu

<sup>11</sup>Rovina  $p$  je komplexní rovina, ve které leží obrazy v Laplaceové transformaci. Používá se pro ni také označení rovina  $s$ .

polopřímky vycházející z počátku v rovině  $z$ . Toto zobrazení je nelineární. Odvodíme je ze vztahu (4.76), který přepíšeme pro imaginární osu roviny  $p$  a jednotkovou kružnici v rovině  $z$ :

$$j\omega_A = \frac{2}{T} \frac{1 - e^{-j\theta}}{1 + e^{-j\theta}} = \frac{2}{T} \frac{2j \sin(\theta/2)}{2 \cos(\theta/2)} = \frac{2j}{T} \tan \frac{\theta}{2}.$$

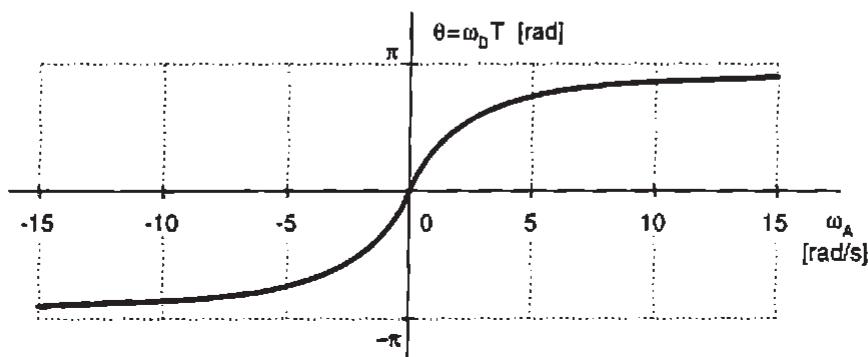
Odtud

$$\omega_A = \frac{2}{T} \tan \frac{\theta}{2}, \quad (4.78)$$

$$\theta = 2 \arctan \frac{\omega_A T}{2}, \quad (4.79)$$

$$\omega_D = \frac{2}{T} \arctan \frac{\omega_A T}{2}, \quad (4.80)$$

kde  $\omega_A$  je frekvence analogového filtru a  $\theta$  resp.  $\omega_D$  jsou odpovídající frekvence filtru číslicového. Vztah (4.79) znázorňuje přibližně obr. 4.28.



Obrázek 4.28 Souvislost mezi frekvencemi analogového a číslicového filtru při bilineární transformaci

Z obr. 4.28 je zřejmé, že u filtru získaného bilineární transformací nemůže dojít k aliasingu, ale mezní frekvence analogového a číslicového filtru nejsou totožné. Také fázová charakteristika se při transformaci deformeuje.

Bilineární transformace zachovává základní vlastnosti vzorových analogových filtrů (monotonost a zakmitávání frekvenční odezvy) a řád filtru. Nezachovává ale tvar impulsní odezvy ani odezvy na skok. Požaduje-li se zachování impulsní odezvy, je nutno použít jiný transformační vztah (metodu invariantnosti impulsní odezvy, viz např. [68, 15]).

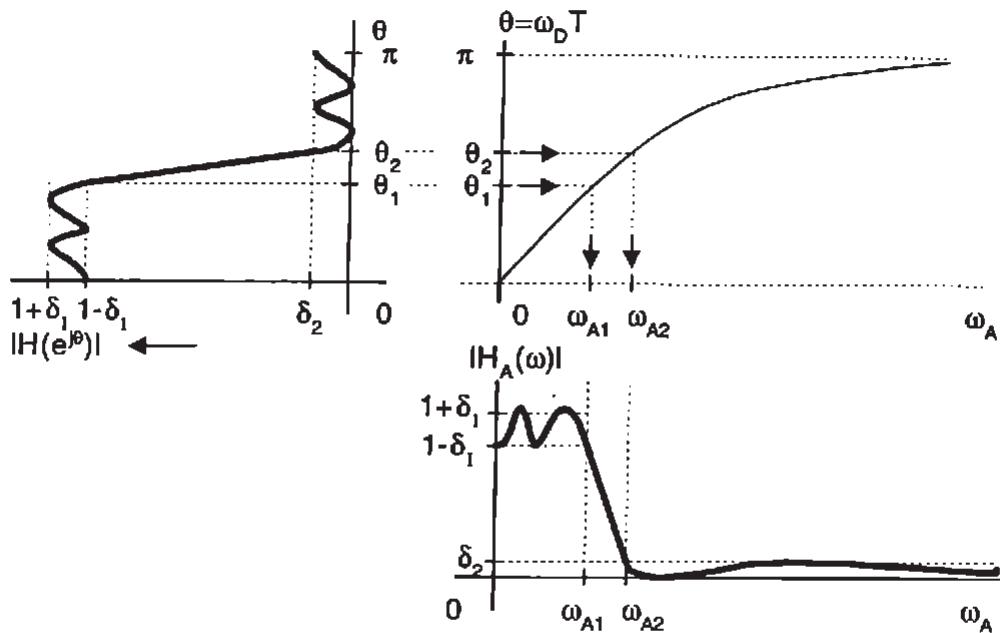
**Příklad 4.3** Nalezení frekvenční charakteristiky číslicové dolní propusti z analogové dolní propusti. Mezní frekvence přechodného pásma číslicové dolní propusti mají být  $f_1 = 2\text{kHz}$ ,  $f_2 = 2,6\text{kHz}$ , vzorkovací frekvence je  $f_{vz} = 8\text{kHz}$  (hodnota používaná v telekomunikacích). ▲

**Řešení** K transformaci charakteristiky použijeme obr. 4.29. Chceme-li navrhnout diskrétní filtr s mezními frekvencemi  $\theta_1, \theta_2$ , musíme nejprve vypočítat mezní frekvence vzorového analogového filtru tak, aby byl splněn vztah (4.79). Pro zadané hodnoty platí

$$T = 125\mu\text{s}, \quad \theta_1 = 0,5\pi, \quad \theta_2 = 0,65\pi,$$

a pro mezní frekvence analogového filtru platí

$$\omega_{A1} = 2\pi f_1 = \frac{2}{T} \tan \frac{\theta_1}{2} = 16000\text{rad/s}, \quad \omega_{A2} = 26100\text{rad/s}.$$



Obrázek 4.29 Souvislost mezi frekvencemi analogové a číslicové dolní propusti; šipky značí postup nalezení frekvencí

**Příklad 4.4** Použijte bilineární transformace k nalezení přenosu  $H(z)$  filtru odpovídajícího pasivnímu integračnímu článku.

**Řešení** Platí

$$\begin{aligned} H_A(p) &= \frac{1}{1+p\tau} = \frac{1/\tau}{1/\tau + p}, \\ H(z) &= \frac{1/\tau}{\frac{1}{\tau} + \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}} = \frac{\frac{T}{T+2\tau}(1+z^{-1})}{1 + (\frac{T-2\tau}{T+2\tau})z^{-1}}. \end{aligned}$$

### Frekvenční transformace

Frekvenčně selektivní filtry (DP, HP, PP a PZ) lze získat z prototypové dolní propusti transformacemi pomocí racionálních lomených funkcí. Komplexní proměnnou charakterizující prototypovou dolní propust značíme  $z^*$  a podobně označíme hvězdičkou i koeficienty přenosu prototypové dolní propusti. Přenos prototypové DP v kaskádní formě tedy je

$$H_{DP}(z^*) = \prod_{i=1}^P \frac{b_{0i}^* + b_{1i}^* z^{*-1} + b_{2i}^* z^{*-2}}{1 - a_{1i}^* z^{*-1} - a_{2i}^* z^{*-2}}. \quad (4.81)$$

Přenos požadovaného frekvenčně selektivního filtru označíme  $H_d(z)$ . Pro přechod od proměnné  $z^*$  k  $z$  bude platit vztah

$$z^{*-1} = G(z^{-1}) \quad (4.82)$$

takový, že

$$H_d(z) = H_{DP}[G^{-1}(z^{*-1})].$$

Aby kauzální a stabilní DP a přenosem  $H_{DP}(z^*)$  přešla do kauzálního a stabilního filtru s přenosem  $H_d(z)$ , musí (4.82) být racionální lomená funkce zobrazující vnitřek jednotkového kruhu v rovině  $z^*$  na vnitřek (nebo jeho část) jednotkového kruhu v rovině  $z$ . Takové funkce jsou publikovány pro přechody  $DP^* \rightarrow DP$ ,  $DP^* \rightarrow HP$ ,  $DP^* \rightarrow PP$  a  $DP^* \rightarrow PZ$  např. v [68, 85]. Protože nepředpokládáme, že by dnes někdo tyto přepočty dělal bez počítače, příslušné vztahy zde neuvádíme.

### 4.3 Transpozice číslicových filtrů

Změnu struktury filtru při zachování jeho přenosu  $H(z)$  umožňuje

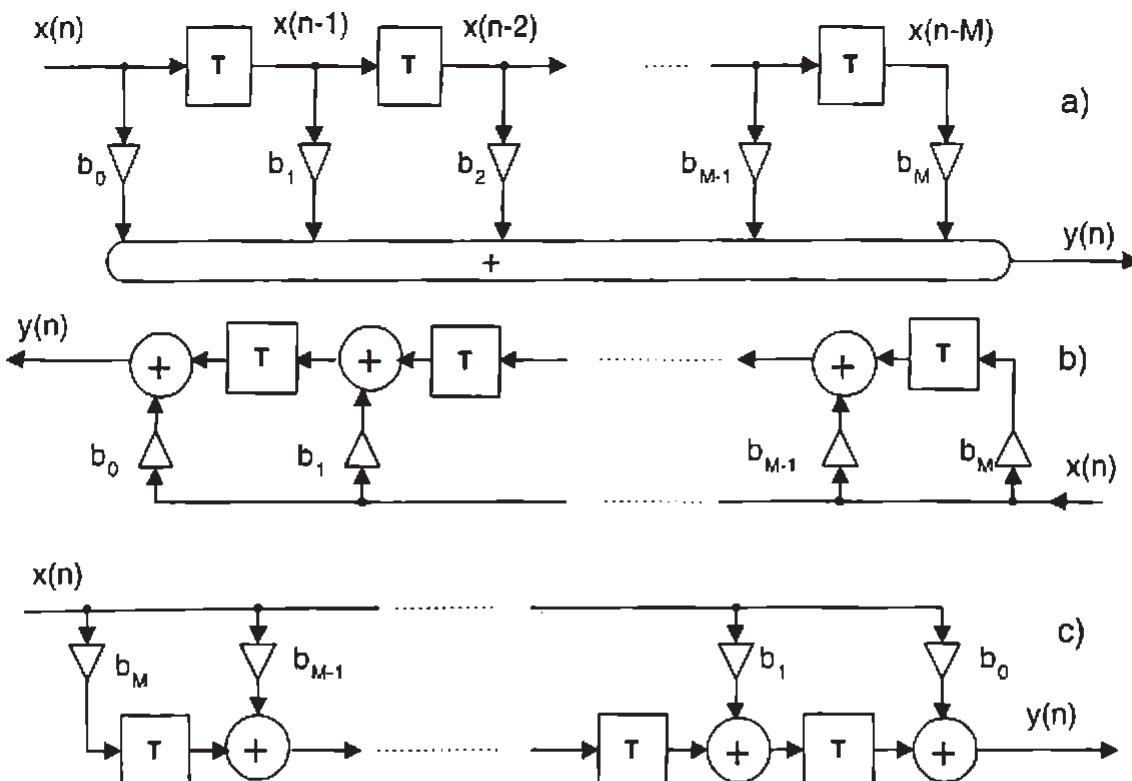
**Věta 4.1 O transpozici.** *Přenos lineárního časově invariantního systému se nezmění, jestliže obrátíme směr toku signálu a vzájemně zaměníme uzly a sčítacky.* ▲

Struktury před a po transpozici se nazývají *duální struktury*.

Duální struktura může být výhodnější z hlediska implementace na určitém druhu integrovaného obvodu. Modifikovaný transverzální filtr dle obr. 4.30 c je např. základem signálového procesoru IMS A100 firmy INMOS, který lze výhodně používat v kaskádném zapojení.

**Příklad 4.5** Najděte pomocí transpozice duální strukturu transverzálního FIR filtru řádu  $M$ . ▲

**Řešení** Uvádí obr. 4.30. ■



**Obrázek 4.30** Transpozice transverzálního FIR filtru; a) původní filtr, b) duální filtr, c) úprava otočením schématu

## 4.4 Porovnání vlastností FIR filtrů a IIR filtrů

Volba typu filtru není jednoznačná a závisí na aplikaci. Výhody FIR filtrů jsou nevýhodami IIR filtrů naopak.

*Filtry FIR:*

- jsou vždy stabilní,
- fázová charakteristika může být (a v praxi vždy je) přesně lineární,
- řád filtru bývá vysoký, proto se návrh dělá pomocí počítače (iteračně), zpoždění filtru je poměrně velké a při hardwarové realizaci je nutný velký počet obvodů,
- nejužívanější je transverzální struktura, která je vhodná i pro realizaci adaptivních filtrů,
- případný rušivý impuls ovlivní výstup jen krátkou dobu,
- vliv chyb kvantování závisí na řádu filtru; nemůže ale dojít k nestabilitě a vzniku mezních cyklů (nelineárních kmitů na výstupu).

*Filtry IIR:*

- mohou být nestabilní,
- používají se převážně jako základní frekvenčně selektivní filtry (DP, HP, PP a PZ); obvykle je zadán pouze průběh amplitudové frekvenční charakteristiky, fázi je možno upravit fázovým posouvačem v kaskádě s daným filtrem,
- návrh je jednodušší, řád filtru je nízký (do 10) a tedy rychlosť reakce vyšší než u FIR filtru,
- i krátký rušivý impuls na vstupu může trvale ovlivnit výstup,
- nejužívanější struktura je kaskáda členů druhého řádu,
- v důsledku chyb kvantování může dojít k nestabilitě (přesune-li se některý z pólů přenosu vně jednotkového kruhu) a ke vzniku mezních cyklů (trvalých kmitů na výstupu při konstantním vstupu).

## 5. Zpracování stochastických signálů

Vedle *deterministických signálů*, které lze popsat matematickou funkcí času a určit tak jejich hodnotu v libovolném okamžiku, představují ve vědě a technice důležitou skupinu *signály náhodné (stochastické)*. Jejich hodnotu pro určitý okamžik v budoucnu nelze vypočítat, ani když máme záznam jejich celého průběhu až do současného okamžiku (tzv. *realizaci*). Tyto signály lze popisovat pouze statisticky.

*Stochastický signál (stochastický proces)* je náhodná funkce času. V určitém okamžiku  $t_0$  je hodnota této náhodné funkce náhodná veličina, která se nazývá *řez náhodným procesem*. Stochastický proces lze tedy definovat jako množinu těchto náhodných veličin pro všechna reálná  $t$  pro analogové procesy, případně pro všechna diskrétní  $t_n = n \cdot T$ ,  $n$ -celé, pro procesy diskrétní v čase. Jinak je možno *náhodný proces* chápat jako nekonečný soubor realizací tohoto procesu. Pro procesy spojité v čase jsou těmito realizacemi funkce času, pro procesy diskrétní v čase jde o posloupnosti. Pojmy náhodný signál a náhodný proces se používají jako synonyma.

Příklady realizací náhodných signálů jsou záznam šumu elektrických obvodů, elektrických projevů živých organismů, výšky mořské hladiny v určitém místě, tloušťky vyráběného papíru na výrobní lince apod.

I náhodné signály mohou být spojité či nespojité v čase a v amplitudě (viz obr. 1.1). Zvláštním typem jsou signály impulsové, jaké získáme např. na výstupu Geiger-Müllerovy trubice. Jde o opakující se impulsy, jejichž okamžiky výskytu jsou náhodné. Pro zpracování diskrétních stochastických signálů využíváme algoritmy popsané v předchozích dvou kapitolách.

Každý signál v praxi lze považovat za součet deterministického a stochastického signálu. Často je jedna z těchto složek zanedbatelná, ale v řadě případů tomu tak není. Složky, která nenesou pro nás užitečnou informaci, se pak snažíme vhodnými metodami (filtrací) zbavit. Příkladem je měření deterministického signálu za přítomnosti šumu.

Pro měření stochastických signálů se používají speciální přístroje (dnes zejména analyzátoru signálu a spektrální analyzátoru, v nedávné minulosti také korelátoru, statistické analyzátoru, měřiče výkonové spektrální hustoty). Moderní přístroje obsahují počítač a vzorkovací jednotky na vstupech. S podstatně menšími náklady (a omezenějšími rozsahy a přesnostmi) lze stochastické signály analyzovat pomocí osobních počítačů se zásuvnými kartami pro digitalizaci vstupních dat.

Pro testování přístrojů a identifikaci objektů je důležité také *vytváření náhodných signálů se zadanými vlastnostmi*. Generování signálů s rovnoměrným rozdělením pravděpodobnosti je běžnou součástí aplikačních programů. Ze signálu s rovnoměrným rozdělením pravděpodobnosti lze různými metodami generovat signály s jinými rozděleními. Nejčastěji se z něj generuje signál s normálním rozložením amplitud. Náročnější je generování signálů se zadanými průběhy jak hustoty pravděpodobnosti, tak autokorelační funkce.

Stochastické signály se dělí na dvě základní skupiny, signály stacionární a signály nestacionární.

*Stacionární stochastický signál* je signál, jehož statistické vlastnosti se nemění s časem. V praxi se ověřuje nezávislost na čase pro střední hodnotu (viz str. 81) a dispersi (str. 82) a

to, zda autokorelační funkce (str. 86) závisí pouze na časovém posuvu a nikoliv na okamžiku počátku měření. Musí tedy platit [84]

$$\mu_x(t_1) = E[x(t_1)] \neq f(t_1), \quad R(t_1, \tau) = E[x(t_1)x(t_1 + \tau)] \neq f(t_1). \quad (5.1)$$

Signály vyhovující tomuto ověření jsou *slabě stacionární* čili *stacionární v širším smyslu*. Signály, u kterých na čase nezávisí hustoty pravděpodobnosti všech řádů, se nazývají *silně stacionární* resp. *striktně stacionární*, *stacionární v užším smyslu*. Splnění tohoto kritéria je prakticky neověřitelné a v praxi se silná stacionárnost předpokládá i u signálů slabě stacionárních. (To je zdůvodnitelné skutečností, že mnoho stochastických signálů vzniká v důsledku působení řady více méně nezávislých vlivů s malými váhami a podle centrální limitní věty jde tedy o gaussovské procesy. Protože gaussovský proces je plně popsán středními hodnotami a kovariancemi, implikuje u něj slabá stacionarita silnou stacionaritu.)

Ve skutečnosti jsou všechny reálné signály nestacionární. Za stacionární považujeme signály, jejichž statistické vlastnosti jsou prakticky neproměnné po dostatečně dlouhou dobu. Délka této „dostatečně dlouhé“ doby závisí na konkrétní aplikaci. Typickými nestacionárními signály jsou řeč a hudba (signál z mikrofonu).

*Ergodické náhodné signály* jsou takové, jejichž statistické parametry (např. střední hodnota) vypočtené z jedné realizace („časové průměry“) se rovnají odpovídajícím parametry vypočteným ze souboru realizací („pravděpodobnostní průměry, průměry v souboru“). Rovná se tedy např. střední hodnota v čase jedné realizace střední hodnotě náhodné veličiny, která je řezem náhodného procesu v určitém časovém okamžiku. Pro charakterizování ergodických náhodných procesů stačí jediná realizace. V praxi se ergodičnost u měřených signálů většinou předpokládá.

Nestacionární náhodné procesy mohou mít v čase proměnnou např. střední hodnotu, rozptyl nebo frekvenční spektrum (viz str. 16). Patří k nim také přechodné náhodné procesy, např. otřesy při zemětřesení nebo explozi.

Zvášťu skupinou signálů jsou *signály pseudonáhodné*. Ty jsou sice vytvářeny nenáhodně, ale svými vlastnostmi se podobají signálům náhodným. Výhodou je jejich snadné generování a reprodukovatelnost.<sup>1</sup> Jsou sice periodické, ale jejich perioda je tak veliká, že z hlediska jejich aplikace se jeví jako signály stochastické.

Náhodné signály se popisují z hlediska *amplitudové struktury* a z hlediska *frekvenční nebo časové struktury*. Popisy ve frekvenční a časové oblasti jsou vzájemně zastupitelné (mezi těmito oblastmi je pro stacionární signály vzájemně jednoznačný přechod). Popis z hlediska amplitudové struktury (hustoty pravděpodobnosti amplitud) nesouvisí v obecném případě s popisem v časové a frekvenční oblasti - náhodné signály se stejnou hustotou pravděpodobnosti amplitud mohou mít různé časové průběhy a různé autokorelační funkce a výkonové spektrální hustoty.

K popisu náhodných signálů můžeme použít:

- funkční závislosti, které mohou popisovat signál
  - z hlediska amplitudové struktury (hustota pravděpodobnosti  $f(x)$  a distribuční funkce  $F(x)$ ),
  - z hlediska časového průběhu (v časové oblasti - korelační a kovarianční funkce, nebo ve frekvenční oblasti - výkonová spektrální hustota),
- číselné charakteristiky - obecné a centrovány momenty, např. střední hodnotu a rozptyl.

<sup>1</sup>Pseudonáhodnými signály jsou i signály s rovnoměrným a normálním rozdělením pravděpodobností generované v počítači různými procedurami s názvy jako rand, randn apod.

## 5.1 Popis náhodných signálů z hlediska amplitudové struktury

### 5.1.1 Popis amplitudového rozdělení pomocí funkčních závislostí

K tomuto popisu používáme distribuční funkce  $F(x)$  a histogram, u amplitudově spojitéch signálů také hustotu pravděpodobnosti  $f(x)$ .

*Distribuční funkce  $F(u)$*  náhodné veličiny  $U$  je definována vztahem

$$F(u) = P(U \leq u), \quad (5.2)$$

kde  $P(U \leq u)$  je pravděpodobnost toho, že veličina  $U$  nabude hodnoty  $\leq u$ . Je-li  $U$  diskrétní náhodná veličina, pak může nabývat jen určitých hodnot  $u_1, u_2, \dots, u_N$ , a to s pravděpodobnostmi  $P_1, P_2, \dots, P_N$ . Platí pro ni vztah

$$F(u_i) = \sum_{U \leq u_i} P_i. \quad (5.3)$$

*Spojitá náhodná veličina* má spojitu distribuční funkci. Distribuční funkce je neklesající a platí pro ni

$$F(-\infty) = 0, \quad F(\infty) = 1. \quad (5.4)$$

Spojité náhodné veličiny lze popsát *hustotou pravděpodobnosti*  $f(u)$ , pro kterou platí

$$f(u) = \frac{dF(u)}{du}. \quad (5.5)$$

V řadě aplikací se vyskytuje *simultánní rozdělení pravděpodobností* soustavy dvou nebo více náhodných veličin  $U_1, U_2, \dots, U_N$ . *Dvourozměrná distribuční funkce* se definuje vztahem

$$F(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2) \quad (5.6)$$

a *dvourozměrná hustota pravděpodobnosti* (pro spojité náhodné veličiny) se definuje vztahem

$$f(u_1, u_2) = \frac{\partial^2 F(u_1, u_2)}{\partial u_1 \partial u_2}. \quad (5.7)$$

Pro statisticky nezávislé veličiny platí

$$F(u_1, u_2) = F(u_1) \cdot F(u_2), \quad (5.8)$$

$$f(u_1, u_2) = f(u_1) \cdot f(u_2). \quad (5.9)$$

Poznatky uvedené pro náhodné veličiny platí obdobně pro náhodné funkce  $x(t)$ . Soubor hodnot náhodného procesu  $x(t)$  v určitém okamžiku  $t = t_0$  (*řez náhodným procesem*) je náhodná veličina. Její rozdělení charakterizuje *distribuční funkce prvního řádu*  $F(x, t)$ , resp.  $F_1(x, t)$ . Pro nestacionární proces závisí na  $t$ , pro stacionární proces je pouze funkce  $x$ . Pro amplitudově spojité procesy můžeme k popisu použít také *hustotu pravděpodobnosti prvního řádu*  $f(x, t)$ , resp.  $f_1(x, t)$ . Souvislost rozdělení hodnot (amplitud) náhodného procesu ve dvou různých časových okamžicích popisují distribuční funkce a hustota pravděpodobnosti druhého řádu. Platí:

$$\begin{aligned} F_2(x_1, t_1; x_2, t_2) &= P(X_1 \leq x_1; X_2 \leq x_2), \\ f_2(x_1, t_1; x_2, t_2) &= \frac{\partial^2 (x_1, t_1; x_2, t_2)}{\partial x_1 \partial x_2}. \end{aligned}$$

## 5.1 Popis náhodných signálů z hlediska amplitudové struktury

Tzv. druhý pravděpodobnostní element náhodného procesu

$$PE_2 = f_2(x_1, t_1; x_2, t_2) \cdot dx_1 dx_2 \quad (5.10)$$

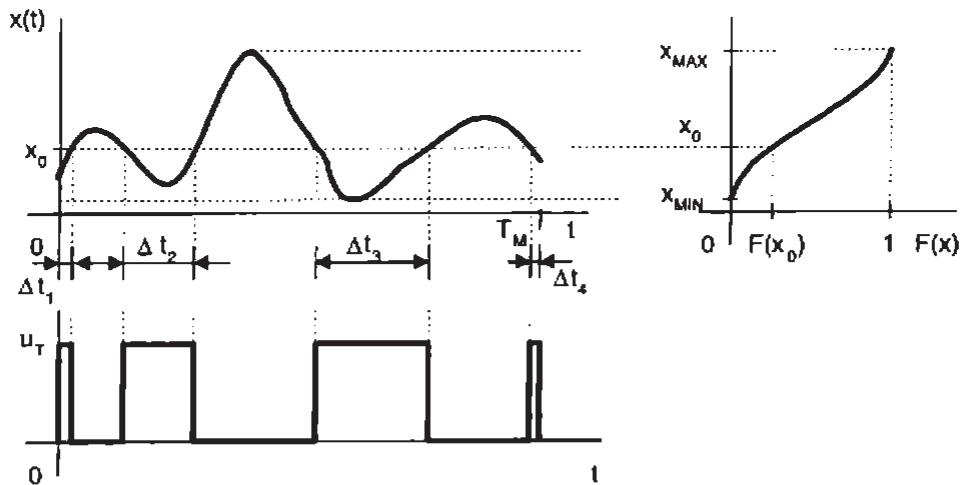
vyjadřuje pravděpodobnost, že realizace náhodného procesu  $x(t)$  je pro  $t = t_1$  v mezích  $x_1$  a  $x_1 + dx_1$  a pro  $t = t_2$  v mezích  $x_2$  a  $x_2 + dx_2$ .

Pojmy distribuční funkce, hustota pravděpodobnosti a pravděpodobnostní element lze rozšířit na libovolný řád  $n$ . Náhodný proces je plně popsán hustotami pravděpodobnosti nebo distribučními funkcemi rádu  $n$  pro libovolné  $n$ .

V praxi se většinou předpokládá ergodičnost náhodného procesu a rozdělení hodnot se určuje z jedné realizace, protože větší počet realizací nebývá k dispozici. Distribuční funkce se určuje podle vztahu

$$F(x_0) = P(x(t) \leq x_0) = \lim_{T_M \rightarrow \infty} \frac{\sum_{i=1}^n \Delta t_i}{T_M}. \quad (5.11)$$

Význam symbolů  $\Delta t_i$  a  $T_M$  i způsob měření  $F(x)$  naznačuje obr. 5.1. V něm je napětí  $u_T$  pomocné napětí nenulové a konstantní během intervalů  $\Delta t_i$ , kdy je  $x(t) \leq x_0$ . Jeho pomocí měříme  $F(x_0)$  dle (5.11), analogově (integrací  $u_T$ ) nebo číslicově (čítačem čítajícím pulsy přes hradlo otevírané signálem  $u_T$ ).



Obrázek 5.1 Princip měření distribuční funkce  $F(x)$  z jedné realizace ergodického náhodného procesu

Protože v praxi měříme vždy konečnou dobu  $T_M$ , získáme pouze odhad distribuční funkce:

$$\hat{F}(x) = \frac{1}{T_M} \sum_{i=1}^n \Delta t_i. \quad (5.12)$$

K proměření celého rozsahu hodnot  $F(x)$  musíme měnit hodnotu  $x_0$  od  $x_{MIN}$  do  $x_{MAX}$ .

Pro amplitudově spojité signály je definována hustota pravděpodobnosti jako derivace distribuční funkce. Pro stacionární signály nezávisí první hustota pravděpodobnosti na čase, index 1 se vypouští a mluví se prostě o hustotě pravděpodobnosti. Platí

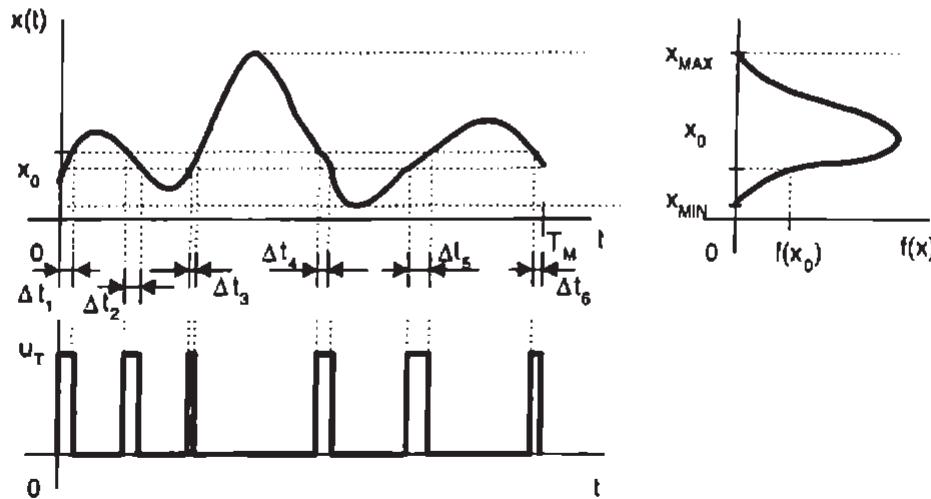
$$f(x) = \frac{dF(x)}{dx} \quad (5.13)$$

a dále

$$f(x) \geq 0, \quad F(x) = \int_{-\infty}^x f(x) dx, \quad (5.14)$$

$$\int_{-\infty}^{\infty} f(x) dx = 1, \quad P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x) dx. \quad (5.15)$$

Princip měření hustoty pravděpodobnosti demonstruje obr. 5.2.



Obrázek 5.2 Princip měření hustoty pravděpodobnosti  $f(x)$  z jedné realizace ergodického náhodného procesu

Hustota pravděpodobnosti se určuje ze vztahu

$$f(x_0) = \lim_{\Delta x \rightarrow 0} P[x_0 < x(t) < x_0 + \Delta x] = \lim_{T_M \rightarrow \infty} \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} \frac{1}{T_M} \sum_{i=1}^n \Delta t_i. \quad (5.16)$$

V praxi je amplitudové okno  $\Delta x$  konečné a měříme po konečnou dobu, takže měřením získáme *odhad hustoty pravděpodobnosti*

$$\hat{f}(x_0) = \frac{\sum_i \Delta t_i}{\Delta x \cdot T_M}. \quad (5.17)$$

Při měření tedy musíme měřit časové intervaly  $\Delta t_i$  a sečít jejich délky. To můžeme udělat analogově integrací napětí  $u_T$  v obr. 5.2, které získáme na výstupu tzv. okénkového komparátoru, porovnávajícího napětí  $x(t)$  s hodnotami  $x_0$  a  $x_0 + \Delta x$ . Číslicový měřič získáme, měříme-li  $\sum \Delta t_i$  a  $T_M$  čítáním impulsů krystalového oscilátoru hradlováných signálem  $u_T$ .

Při měření musíme měnit hodnotu  $x_0$  přes celé pásmo  $x_{MIN}$  až  $x_{MAX}$ , ve kterém se vyskytují hodnoty  $x(t)$ . Zrychlení měření se dosáhne vícekanálovým měřičem, kde je několik okénkových komparátorů srovnávajících signál  $x(t)$  současně s několika hodnotami  $x_{0k} \pm \Delta x_{0k}/2$ . (Z hlediska přesnosti je lepší měřit v pásmu amplitud  $\Delta x$  rozloženém souměrně okolo hodnoty  $x_0$ , než mezi hodnotami  $x_0$  a  $x_0 + \Delta x$ .)

### Výpočet rozdělení amplitud počítačem - histogram

Předpokládáme ergodický náhodný signál  $x(t)$ . Vzorkováním jedné realizace dostaneme posloupnost  $x(n) = x(nT)$ , kde  $T$  je vzorkovací interval. Hodnoty této posloupnosti leží v rozsahu

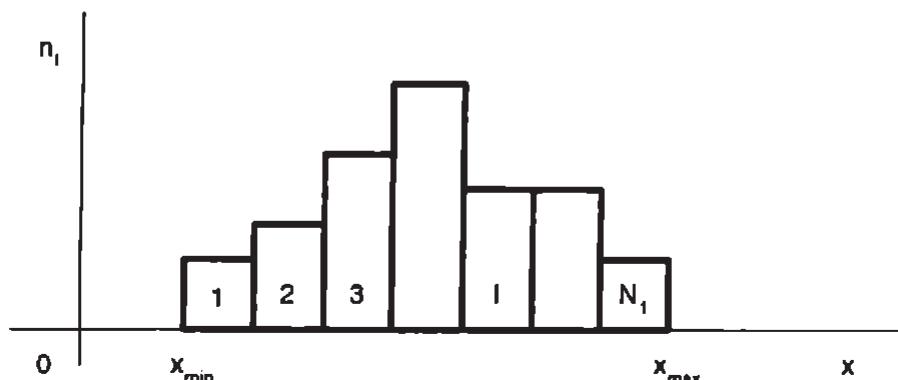
$(x_{MIN}, x_{MAX})$ . Postup nalezení histogramu závisí na tom, známe-li předem všechny vzorky (měření off-line), nebo měříme-li průběžně (on-line).

Známe-li předem všechn  $N$  hodnot posloupnosti vzorků, můžeme tyto hodnoty rozdělit do  $N_1$  intervalů, zvaných někdy *třídní intervaly*. Šířka těchto intervalů se obvykle volí konstantní a určí se jako

$$\Delta x = \frac{x_{MAX} - x_{MIN}}{N_1}. \quad (5.18)$$

Do  $l$ -tého intervalu padne  $n_l$  hodnot. Kontrolu správnosti výpočtu lze provést kontrolou, zda je splněn vztah

$$\sum_{l=1}^{N_1} n_l = N. \quad (5.19)$$



Obrázek 5.3 Histogram absolutních četností signálu  $x(n)$

Pokud na svislou osu vynášíme počty hodnot  $x(n)$ , které padly do  $l$ -tého intervalu (*absolutní četnosti*), jde o histogram absolutních četností (viz obr. 5.3). Vyneseme-li na svislou osu hodnoty  $n_l/N$ , jde o histogram relativních četností.

V případě měření v reálném čase (čili on-line) neznáme hodnoty předem. Proto nemůžeme použít vztah (5.18). Pak zvolíme  $N_1$  a  $\Delta x$  a oba krajní intervaly volíme  $(-\infty, x_a)$  a  $(x_b, +\infty)$ . Do těchto dvou intervalů zařazujeme hodnoty, pro které platí  $x < x_a$  nebo  $x > x_b$ . Intervaly v tomto případě číslujeme  $0, 1, \dots, N_1, N_1 + 1$ .

Rychlý algoritmus výpočtu histogramu je následující postup, který aplikujeme na každou hodnotu  $x(n)$ :

1. Je-li  $x(n) < x_a$ , přidáme jedničku k  $n_0$  a začneme třídit následující hodnotu  $x(n+1)$ ; je-li  $x(n) \geq x_a$ , zkoumá se, je-li  $x(n) \geq x_b$ .
2. Je-li  $x(n) \geq x_b$ , přidá se jednička k  $n_{N+1}$  a přikročí se k třídění  $x(n+1)$ .
3. Vyjde-li z bodů 1 a 2, že je  $x_a \leq x(n) < x_b$ , určí se  $k = INT[(x - x_a)/\Delta x]$  a přidá se jednička k  $n_k$ .

Tvar histogramu pro danou posloupnost hodnot  $x(n)$  není jednoznačný, ale závisí na volbě  $N_1$ . Volba počtu třídních intervalů  $N_1$  vyžaduje určitou zkušenosť a řídí se celkovým počtem hodnot  $N$ . Zvolíme-li  $N_1$  příliš velké, projevují se v histogramu značné výkyvy. Při volbě  $N_1$  příliš malého histogram neukáže detaily. Protože hodnoty  $x(n)$  jsou v praxi kvantované v úrovni (jsou získány z výstupu A/D převodníku), musíme volit dílčí intervaly tak, aby každý obsahoval stejný počet kvantovacích úrovní. Při nesplnění této podmínky vyjde histogram signálu s rovnoměrným rozložením pravděpodobností pravidelně zvlněný.

Histogram relativních četností umožňuje získat *odhad rozdělení hustoty pravděpodobnosti*, vydělíme-li hodnoty  $n_l/N$  šírkou okna (intervalu)  $\Delta x$ , čili nad dílčí intervaly vyneseme sloupce o výšce  $n_l/(N \cdot \Delta x)$ .

### Poznámky k teorii odhadů

Protože při zpracování konečného počtu  $N$  vzorků pracujeme z hlediska statistiky s *výběrovým souborem* čili výběrem ze základního souboru, jsou odhady jednotlivých parametrů také náhodnými veličinami, charakterizovanými střední hodnotou a rozptylem. Odhad parametru  $\alpha$  označíme  $\hat{\alpha}$ .

V praxi se projevují dva typy odchylek  $\alpha$  od  $\hat{\alpha}$ , čili chyb odhadu. Jde o *vychýlení odhadu*  $V$  (angl. *bias*) a *rozptyl odhadu*  $R$  (angl. *variance*). Definují je vztahy

$$V = \alpha - E(\hat{\alpha}), \quad (5.20)$$

$$R = \sigma_{\hat{\alpha}}^2 = E[(\hat{\alpha} - E(\hat{\alpha}))^2]. \quad (5.21)$$

*Nevychýlený odhad* je odhad, pro který je  $V = 0$ . Rozptyl odhadu vypovídá o tom, jak je hustota pravděpodobnosti odhadu  $f(\hat{\alpha})$  rozložena okolo hodnoty  $\alpha$ .

Porovnání kvality odhadů je často komplikováno situací, kdy odhad s malým  $V$  má velké  $R$  a naopak. Proto je výhodné používat *středně kvadratickou chybu* (střední chybu kvadrátu odhadu) definovanou vztahem

$$E [(\hat{\alpha} - \alpha)^2] = \sigma^2 + V^2, \quad (5.22)$$

resp. *relativní středně kvadratickou chybu*

$$\varepsilon = \frac{\sqrt{E [(\hat{\alpha} - \alpha)^2]}}{\alpha} = \sqrt{\varepsilon_V^2 + \varepsilon_r^2}. \quad (5.23)$$

Odhad  $\hat{\alpha}$  je *konzistentní odhad* parametru  $\alpha$ , platí-li ([45])

$$\lim_{N \rightarrow \infty} E [(\hat{\alpha} - \alpha)^2] = 0, \quad (5.24)$$

čili je-li odhad nevychýlený a přitom se jeho rozptyl blíží nule při zvyšování počtu vzorků k nekonečnu. Středně kvadratická chyba konzistentního odhadu se blíží nule pro počet pozorování  $N$  rostoucí nad všechny meze.

Parametrem  $\alpha$  ve výše uvedených vztazích může být střední hodnota posloupnosti  $x(n)$  nebo její rozptyl, ale i funkční závislosti jako hustota pravděpodobnosti, korelační funkce nebo výkonová spektrální hustota (viz str. 95).

Při *vzorkování náhodných signálů* se vzorkovací interval  $T$  volí často tak, aby sousední vzorky byly nekorelované, čili aby  $T$  bylo větší než maximální interval korelace (viz str. 85).

#### 5.1.2 Popis amplitudového rozdělení pomocí číselných hodnot

V některých případech stačí k základní představě o rozložení hodnot náhodného signálu pouze několik hodnot. Je to *momenty prvního řádu*, případně veličiny od nich odvozené. Tyto momenty přitom můžeme nalézt zpracováním souboru hodnot řady realizací v určitém časovém okamžiku  $t_0$ , nebo (v praxi půdstatně častěji) zpracováním jediné realizace. Pro ergodické procesy jsou momenty získané oběma způsoby stejné. (Ergodické procesy jsou stacionární, takže hodnoty jejich momentů nezáleží na volbě okamžiku  $t_0$ .)

*K-tý obecný moment* prvního řádu náhodné veličiny  $X$  je definován vztahem

$$\mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx . \quad (5.25)$$

*K-tý centrální (centrovaný) moment* prvního řádu náhodné veličiny  $X$  je definován

$$\mu_k^c = E[(X - \mu_x)]^k = \int_{-\infty}^{\infty} (x - \mu_x)^k f(x) dx , \quad (5.26)$$

kde symbol  $E$  značí střední hodnotu (očekávanou hodnotu) a  $\mu_x$  je střední hodnota veličiny  $X$ . Náhodná veličina  $X$  je řezem náhodného procesu  $x(t)$  pro  $t = t_0$ .

Vztahy (5.25) a (5.26) platí pro (amplitudově) spojité náhodné veličiny. Pro (amplitudově) diskrétní náhodné veličiny platí

$$\mu_k = E(X^k) = \sum_{i=1}^n x_i^k \cdot P_i , \quad (5.27)$$

$$\mu_k^c = E[(X - \mu_x)]^k = \sum_{i=1}^n (x_i - \mu_x)^k \cdot P_i , \quad (5.28)$$

kde  $P_i$  je pravděpodobnost, že veličina  $X^k$  resp.  $(X - \mu_x)^k$  nabude hodnoty  $x_i^k$  resp.  $(x_i - \mu_x)^k$ .

Při výpočtu centrálních momentů se pracuje s odchylkami hodnot náhodného procesu od střední hodnoty tohoto procesu.

Platí (viz [54, 55])

$$\mu_0^c(x) = \mu_0(x) = 1 , \quad (5.29)$$

$$\mu_1^c(x) = 0, \quad \mu_1(x) = E(X) , \quad (5.30)$$

$$\mu_2^c(x) = \mu_2(x) - [E(X)]^2 , \quad (5.31)$$

$$\mu_3^c(x) = \mu_3(x) - 3\mu_2(x) \cdot E(X) + 2 \cdot [E(X)]^2 , \quad (5.32)$$

$$\mu_4^c(x) = \mu_4(x) - 4\mu_3(x) \cdot E(X) + 6\mu_2(x) [E(X)]^2 - 3 [E(X)]^4 . \quad (5.33)$$

Z číselných charakteristik mají základní význam střední hodnota, střední hodnota kvadrátu a rozptyl.

*Střední hodnota* je první obecný moment  $\mu_1$ . Střední hodnota signálu  $x(t)$  v souboru se značí  $\mu_x$  nebo  $E(X)$ , střední hodnota v čase se značí  $\bar{x}$ . Platí

$$\mu_x = \int_{-\infty}^{\infty} x f(x) dx, \quad \text{resp.} \quad \mu_x = \sum_{i=1}^n x_i P_i . \quad (5.34)$$

Pro střední hodnotu v čase platí pro spojitý signál  $x(t)$

$$\bar{x} = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_{-T_M/2}^{T_M/2} x(t) dt , \quad (5.35)$$

a pro diskrétní (vzorkovaný) signál  $x(n) = x(nT)$  je pro  $n \in (-N, N)$

$$\bar{x} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(nT) . \quad (5.36)$$

Ve skutečnosti ovšem pracujeme s časově omezenými signály, konečným počtem vzorků  $N$ , případně s konečným počtem realizací náhodného procesu (takže náhodná veličina  $X$  je výběrovým souborem). Měřením pak získáme pouze odhad střední hodnoty, např.

$$\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x(nT). \quad (5.37)$$

Lze ukázat, že (5.37) je konzistentním odhadem veličiny  $\mu_x$ .

Protože integrace i sumace jsou lineární operace, platí

$$E(X + Y) = E(X) + E(Y), \quad E(k \cdot X) = k \cdot E(X). \quad (5.38)$$

Střední hodnota je charakteristikou polohy rozdělení na ose  $x$ .

*Střední hodnota kvadrátu* je druhý obecný moment (prvního řádu)  $\mu_2$ . Pro výpočet v soubooru platí (pro spojitou resp. diskrétní veličinu)

$$\mu_2 = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx, \quad \text{resp.} \quad E(X^2) = \sum_{i=1}^n x_i^2 P_i. \quad (5.39)$$

Pro signály  $x(t)$  spojité v čase platí:

$$\overline{x^2} = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_{-T_M/2}^{T_M/2} x^2(t) dt = X_{EF}^2, \quad (5.40)$$

kde  $X_{EF}$  je efektivní hodnota signálu  $x(t)$ .

Pro diskrétní signál  $x(n) = x(nT)$  pro  $n \in \{-N, N\}$  platí

$$\overline{x^2} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x^2(nT), \quad (5.41)$$

resp. pro  $N$  vzorků pro kladné časy

$$\overline{x^2} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} x^2(nT). \quad (5.42)$$

Důležitější než střední hodnota kvadrátu je druhý centrální moment  $\mu_2^c$ , čili tzv. *rozptyl* (*disperse, variance*). Obvykle se značí  $D(X)$  nebo  $var(x)$ . Platí pro něj

$$D(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx, \quad \text{resp.} \quad D(X) = \sum_{i=1}^n (x_i - \mu_x)^2 P_i. \quad (5.43)$$

Při určování z jedné realizace platí pro spojité signály

$$\overline{[x(t) - \bar{x}]^2} = E[x(t) - E(x(t))]^2 = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_{-T_M/2}^{T_M/2} [x(t) - \bar{x}]^2 dt. \quad (5.44)$$

Jde o kvadrát efektivní hodnoty fluktuací signálu  $x(t)$  okolo stejnosměrné složky  $X_0 = \bar{x}$ , čili o výkon fluktuací signálu na odporu  $1\Omega$ .

Pro signály diskrétní v čase platí

$$\overline{[x(nT) - \bar{x}]^2} = E[x(nT) - E(x(nT))]^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} [x(nT) - \bar{x}]^2. \quad (5.45)$$

Pro ergodické signály platí

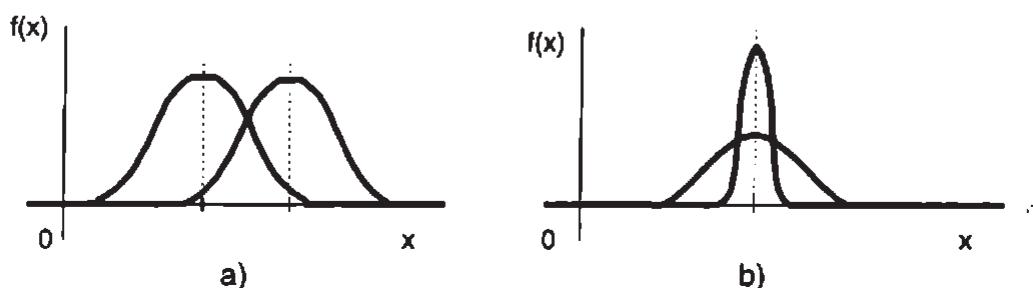
$$D(X) = \overline{[x(nT) - \bar{x}]^2}. \quad (5.46)$$

Pro konečný počet vzorků  $N$  dostaneme *nevychýlený odhad*  $\hat{D}(x)$  ergodického procesu ze vztahu

$$\hat{D}(x) = \frac{1}{N-1} \sum_{n=0}^{N-1} [x(nT) - \hat{\bar{x}}]^2. \quad (5.47)$$

Z (5.39) plyne, že platí také  $E(X^2) = D(X) + \mu_x^2$ . Disperze je mírou rozptýlení hodnot okolo střední hodnoty.

Vliv střední hodnoty a rozptylu na polohu a tvar hustoty pravděpodobnosti amplitud náhodného procesu demonstruje obr. 5.4.



Obrázek 5.4 Hustoty pravděpodobnosti dvou signálů s rozdílnými středními hodnotami a stejnými rozptyly (a) a se stejnými středními hodnotami a rozdílnými rozptyly (b)

*Středně kvadratická odchylka*  $\sigma_x$  (směrodatná odchylka) je definována vztahem

$$\sigma_x = \sqrt{D(X)}. \quad (5.48)$$

Její odhad se počítá ze vztahu

$$s = \hat{\sigma}_x = \sqrt{\hat{D}(X)}. \quad (5.49)$$

Kromě disperze charakterizují *tvor rozdělení* také asymetrie a špičatost. Jejich definice viz např. [54].

Dalšími důležitými parametry jsou medián, modus a P-kvantil. Jde o charakteristiky náhodné veličiny (která je zde chápána jako řez náhodného signálu pro určitý okamžik).

*Medián* (centrální hodnota) leží ve středu posloupnosti hodnot  $x$  seřazených podle velikosti. Při lichém počtu hodnot je roven hodnotě ve středu usporádané posloupnosti. Při sudém počtu hodnot je medián roven aritmetickému průměru hodnot ležících ve středu (např. pro hodnoty 1 2 3 3 4 5 5 6 je medián roven  $(3+4)/2 = 3,5$ ). O mediánu se bude ještě mluvit v části o zpracování obrazu, viz 11.3.2.

*Modus* je hodnota  $x_M$ , pro kterou nabývá hustota pravděpodobnosti spojité náhodné veličiny maxima, případně pro kterou je největší hodnota  $P_i$  diskrétní náhodné veličiny. Platí tedy

$$f(x_M) \geq f(x) \text{ pro } -\infty < x < \infty, \text{ resp.} \quad (5.50)$$

$$P(X = x_M) \geq P(X = x_i) \text{ pro } i = 1, 2, \dots, n. \quad (5.51)$$

$P$ -kvantil náhodné veličiny  $X$  je číslo  $x_P$  takové, že pro dané  $0 \leq P \leq 1$ , platí pro diskrétní náhodné veličiny  $F(x_P) = P(X \leq x_P) \leq P$  resp. pro spojité náhodné veličiny  $F(x_P) = P$ . Kvantity jsou jednoznačně určeny pro distribuční funkce spojité a rostoucí pro libovolné  $x$ .

### 5.1.3 Dvě základní spojité rozdělení pravděpodobnosti

Nejdůležitější spojité rozdělení pravděpodobností jsou normální a rovnoměrné. Zopakujeme zde jejich základní vlastnosti.

#### Normální rozdělení

Normální čili Gaussovo rozdělení pravděpodobnosti je nejdůležitějším spojitym rozdělením. Náhodná veličina  $X$  je rozdělena normálně, působí-li při jejím vzniku řada libovolně rozložených ovlivňujících veličin, které jsou nezávislé a mají přibližně stejnou váhu.

Náhodná veličina  $X$  má normální rozložení  $N(\mu, \sigma)$  s parametry  $\mu$  a  $\sigma$ , ( $-\infty < \mu < \infty, \sigma > 0$ ), platí-li pro její hustotu pravděpodobnosti vztah

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (5.52)$$

Funkce  $f(x)$  je souměrná kolem  $x = \mu$ , což je modus rozdělení a zároveň střední hodnota. Tvar závisí na hodnotě  $\sigma$ . Důležitým zvláštním případem je normované normální rozdělení s parametry  $\mu = 0$ ,  $\sigma = 1$ , které se označuje  $N(0, 1)$ . Toto rozdělení je tabelováno<sup>2</sup> a pomocí jeho tabulek lze určovat pravděpodobnosti i pro obecné normální rozdělení. Dnes jsou ale podobné úlohy řešeny pomocí počítače.

U normálního rozdělení leží přibližně 68,3% hodnot v pásmu  $\mu \pm \sigma$ , 95,5% hodnot v pásmu  $\mu \pm 2\sigma$  a 99,7% hodnot v pásmu  $\mu \pm 3\sigma$ . Ačkoliv tedy teoreticky může  $x$  nabýt libovolné hodnoty, 99,7% těchto hodnot leží v pásmu  $\pm 3\sigma$  okolo střední hodnoty. Protože se při měření v praxi většinou předpokládá (na základě centrální limitní věty teorie pravděpodobnosti), že jsou náhodné chyby normálně rozložené, bere se jako krajní chyba právě hodnota  $3\sigma$ . Pro odhad hodnoty  $\sigma$  se používá vztahů (5.47) a (5.49). Pro normální rozdělení platí

$$\text{var}(X) = \sigma^2, \quad E(X) = \mu. \quad (5.53)$$

Důležité je, že lineární funkce nezávislých normálně rozložených veličin  $X_i$ ,  $Y = \sum_{i=1}^N a_i X_i + b$  je náhodná veličina rozdělená normálně. Platí pro ni

$$E(Y) = \sum_{i=1}^N a_i \mu_i + b, \quad D(Y) = \sum_{i=1}^N a_i^2 \sigma_i^2. \quad (5.54)$$

Aplikujeme-li tedy na posloupnost normálně rozložených čísel lineární operaci (např. FFT, FTD nebo číslicovou filtraci), jsou výsledné hodnoty opět normálně rozloženy. To platí také pro aritmetický průměr  $\bar{X}$ .  $N$  vzájemně nezávislých normálně rozdělených veličin veličin  $X_i$ . Mají-li veličiny  $X_i$  stejné střední hodnoty  $\mu$  a stejné směrodatné odchylky  $\sigma$ , má aritmetický průměr rozdělení  $N(\mu, \sigma/\sqrt{N})$ . Jeho střední hodnota je tedy stejná jako u vstupních veličin, ale rozptyl je  $N$ -krát menší. Toho se využívá při opakování měření konstantní hodnoty. S rostoucím počtem měření se zmenšuje odchylka  $\bar{X}$  od  $\mu$ .

<sup>2</sup>V MATLABu se posloupnost čísel rozdělených podle tohoto rozdělení vyvolá příkazem `randn(N, 1)`.

### Rovnoměrné rozdělení

Toto rozdělení je důležité v měření a řízení proto, že je podle něj rozložena kvantovací chyba A/D převodníků. Kromě toho jsou posloupnosti rovnoměrně rozložených náhodných čísel běžně dostupné na počítačích a lze z nich snadno generovat normálně rozložená čísla nebo čísla s jinými rozloženími pravděpodobnosti.

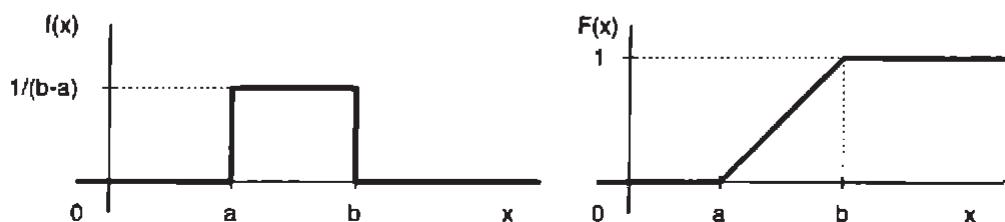
Náhodná veličina  $X$  je rovnoměrně rozložena v intervalu  $(a, b)$ , platí-li pro její hustotu pravděpodobnosti vztah

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a, b), \\ 0 & \text{jinde.} \end{cases} \quad (5.55)$$

Grafy hustoty pravděpodobnosti  $f(x)$  a distribuční funkce  $F(x)$  tohoto rozdělení uvádí obr. 5.5.

Pro  $F(x)$  platí

$$F(x) = \int_{z=-\infty}^x f(z) dz = \frac{x-a}{b-a} \cdot x, \quad a \leq x \leq b. \quad (5.56)$$



Obrázek 5.5 Hustota pravděpodobnosti  $f(x)$  a distribuční funkce  $F(x)$  rovnoměrného rozložení

Dosazením za  $f(x)$  do (5.34) a do (5.43) dostaneme

$$\mu_x = \frac{a+b}{2}, \quad D(X) = \sigma^2 = \frac{(b-a)^2}{12}. \quad (5.57)$$

Důležitým zvláštním případem jsou čísla rovnoměrně rozložená v intervalu  $(0, 1)$ . Pro ně je  $\mu_{x1} = 0,5$  a  $\sigma_{x1} = 1/\sqrt{12}$ .<sup>3</sup> U A/D převodníku s kvantovacím krokem  $q$  je rozptyl kvantovacího šumu roven  $q^2/12$ .

## 5.2 Popis náhodných signálů v časové oblasti

K popisu náhodných signálů v časové oblasti se používají *korelační funkce a kovarianční funkce*. Tyto funkce nesou informaci o souvislosti hodnoty náhodného procesu  $x(t)$  pro  $t = t_0$  s hodnotou tohoto procesu (resp. jiného procesu  $y(t)$ ) pro  $t = t_1 = t_0 + \tau$ . Vyjadřují tedy souvislost mezi dvěma řezy náhodného procesu.

Funkce popisující souvislost hodnot téhož procesu se nazývají *autokorelační* resp. *auto-kovarianční*. Statistickou závislost hodnot dvou různých procesů popisují *vzájemná korelační funkce a vzájemná kovarianční funkce*.

Uvedené funkce jsou *momenty druhého rádu* a jako takové mohou být počítány jednak jako příslušná střední hodnota ze souboru realizací, jednak jako příslušná střední hodnota v čase.

<sup>3</sup>V MATLABu jsou tato čísla generována příkazem `rand`.

Korelační funkce budeme značit  $R$ , kovarianční funkce  $K$ . Pomocí indexů vyjadříme, zda jde o funkci popisující jeden proces  $(R_{xx}, K_{xx})$  nebo dva procesy  $(R_{xy}, K_{xy})$ . Funkce získané zpracováním souboru označíme čárkou, funkce vypočtené jako střední hodnoty v čase budou bez čárky. (Je tedy  $K'_{yx}$  vzájemná kovarianční funkce procesů  $y(t)$  a  $x(t)$ , vypočtená ze souboru realizací.)

### 5.2.1 Korelační a kovarianční funkce určené ze souboru

*Autokorelační funkce* (určená pro okamžiky  $t$  a  $t' = t + \tau$ ) je

$$R'_{xx}(t, t + \tau) = E[x(t) \cdot x(t + \tau)] . \quad (5.58)$$

Pro (amplitudově) spojitý náhodný proces platí

$$R'_{xx}(t, t + \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot x' \cdot f_2(x, t; x', t + \tau) dx dx' . \quad (5.59)$$

V (5.59) je  $f_2(x, t; x', t + \tau)$  druhá hustota pravděpodobnosti procesů  $x(t)$  a  $x'(t) = x(t + \tau)$ . *Autokovarianční funkce*  $K'_{xx}$  je autokorelační funkce centrovaného náhodného procesu:

$$K'_{xx}(t, t + \tau) = E \{ [x(t) - \mu_x(t)] \cdot [x(t + \tau) - \mu_x(t + \tau)] \} . \quad (5.60)$$

Pro stacionární náhodné procesy závisí  $R'_{xx}$  a  $K'_{xx}$  pouze na vzdálenosti obou časových okamžiků  $\tau$ , nikoliv na poloze prvního okamžiku  $t$ . Hodnota  $\mu_x$  je konstantní (nezávisí na čase) a je tedy

$$R'_{xx}(\tau) = E[x(0) \cdot x(\tau)] , \quad (5.61)$$

$$K'_{xx}(\tau) = E \{ [x(0) - \mu_x] \cdot [x(\tau) - \mu_x] \} . \quad (5.62)$$

Platí

$$R'_{xx}(\tau) = K'_{xx}(\tau) + \mu_x^2 . \quad (5.63)$$

*Normovaná autokovarianční funkce* nabývá pro  $t = 0$  hodnotu 1 a pro stacionární proces platí

$$\varrho'_{xx}(\tau) = \frac{R'_{xx}(\tau) - \mu_x^2}{\sigma_x^2} = \frac{K'_{xx}(\tau)}{K'_{xx}(0)} . \quad (5.64)$$

Pro  $\tau = 0$  je totiž  $K'_{xx}(0) = \sigma_x^2 = D(X)$ .

Disperzi procesu čili výkon fluktuací signálu  $x(t)$  kolem  $\mu_x$  na odporu  $1\Omega$  (viz str. 82) lze tedy určit z autokovarianční funkce pro  $\tau = 0$ .

Pro *autokorelační funkci reálných stacionárních signálů* platí

$$R'_{xx}(\tau) = R'_{xx}(-\tau), \quad (\text{sudost autokorelační funkce}), \quad (5.65)$$

$$R'_{xx}(0) > |R'_{xx}(\tau)|, \quad R'_{xx}(0) > 0 , \quad (5.66)$$

$$R'_{xx}(\pm\infty) = \mu_x^2 . \quad (5.67)$$

Odtud (viz (5.64))

$$K'_{xx}(\pm\infty) = 0 . \quad (5.68)$$

Hodnota  $R_{xx}(0)$  je celkový výkon signálu na odporu  $1\Omega$  (včetně výkonu stejnosměrné složky).

Obdobně můžeme určovat ze souboru realizací i vzájemnou korelační a vzájemnou kovarianční funkci. Platí

$$R'_{xy}(t, t + \tau) = E[x(t) \cdot y(t + \tau)], \quad (5.69)$$

$$K'_{xy}(t, t + \tau) = E\{[x(t) - \mu_x(t)] \cdot [y(t + \tau) - \mu_y(t + \tau)]\}, \quad (5.70)$$

$$\varrho'_{xy}(\tau) = \frac{K'_{xy}(t, t + \tau)}{\sigma_x(t) \cdot \sigma_y(t + \tau)}. \quad (5.71)$$

Pro stacionární a reálné náhodné procesy závisí tyto funkce opět pouze na  $\tau$  a platí

$$R'_{xy}(\tau) = R'_{yx}(\tau) \quad (5.72)$$

$$R'_{xx}(0) \cdot R'_{yy}(0) \geq |R'_{xy}(\tau)|^2. \quad (5.73)$$

Jsou-li dva procesy  $x(t)$  a  $y(t)$  nekorelované, platí pro ně pro libovolné  $\tau$

$$K'_{xy}(\tau) = 0, \quad \text{resp.} \quad R'_{xy}(\tau) = \mu_x \mu_y. \quad (5.74)$$

U gaussovských procesů (procesů s normálním rozložením amplitud) plyně z nekorelovanosti i statistická nezávislost.

### 5.2.2 Korelační a kovarianční funkce určené jako střední hodnoty v čase

Pro autokorelační funkci stacionárního náhodného procesu  $x(t)$  definovanou jako střední hodnota v čase (moment druhého rádu v čase) platí

$$R_{xx}(\tau) = \overline{x(t) \cdot x(t + \tau)} = E[x(t) \cdot x(t - \tau)]. \quad (5.75)$$

Pro stacionární náhodný proces spojitý v čase tedy

$$R_{xx}(\tau) = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_0^{T_M} x(t)x(t + \tau) dt, \quad (5.76)$$

$$K_{xx}(\tau) = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_0^{T_M} [x(t) - \mu_x][x(t + \tau) - \mu_x] dt. \quad (5.77)$$

Pro stacionární náhodný proces diskrétní v čase (vzorkovaný) platí

$$R_{xx}(rT) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x(nT)x(nT + rT), \quad (5.78)$$

(časový posuv je zde obvykle celistvým násobkem vzorkovacího intervalu  $T$ ).

Měření probíhá konečnou dobu  $T_M$  resp. zpracovává se konečný počet vzorků  $N$ , takže dostaneme pouze odhady. Výrazy pro tyto odhady můžeme dostat z (5.76) až (5.78), vypustíme-li v nich limity. Tyto odhady jsou ale vychýlené, proto se v praxi pracuje s *odhady autokorelačních funkcí*

$$\hat{R}_{xx}(\tau) = \frac{1}{T_M - \tau} \int_0^{T_M - \tau} x(t)x(t + \tau) dt, \quad (5.79)$$

$$\hat{R}_{xx}(rT) = \frac{1}{N - r} \sum_{n=1}^{N-r} x(nT)x(nT + rT), \quad (5.80)$$

kde největší hodnota  $\tau$  resp.  $\tau T$  je v praxi většinou desetina  $T_M$  resp.  $NT$ .

Výrazy obdobné výrazům (5.75) až (5.80) lze napsat také pro vzájemné korelační a vzájemné kovarianční funkce (nahradíme  $x(t + \tau)$  funkcí  $y(t + \tau)$ ).

Výrazy (5.63) až (5.74) platí také pro korelační a kovarianční funkce určené jako střední hodnoty v čase, tedy nečárkováné.

Pro ergodické náhodné procesy platí

$$R'_{xx}(\tau) = R_{xx}(\tau); \quad K'_{xx}(\tau) = K_{xx}(\tau); \quad \varrho'_{xx}(\tau) = \varrho'_{xx}(\tau), \quad (5.81)$$

$$R'_{xy}(\tau) = R_{xy}(\tau); \quad K'_{xy}(\tau) = K_{xy}(\tau); \quad \varrho'_{xy}(\tau) = \varrho'_{xy}(\tau). \quad (5.82)$$

Důležitou hodnotou, která charakterizuje, jak moc časově vzdálené hodnoty procesu lze už považovat za prakticky nekorelované, je *maximální interval korelace*  $\tau_{km}$ . Určuje se z normované autokovarianční funkce a platí pro něj

$$\varrho_{xx}(\tau_{km}) = 0.05; \quad \varrho_{xx}(\tau) < 0.05 \quad \text{pro } \tau > \tau_{km}. \quad (5.83)$$

Korelační funkce jsou vázány s výkonovými spektrálními hustotami *Wienerovými-Chinčinovými vztahy*

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(f) e^{j2\pi f\tau} df, \quad (5.84)$$

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} S_{xy}(f) e^{j2\pi f\tau} df. \quad (5.85)$$

Tyto vztahy jsou zpětnou Fourierovou transformací.<sup>4</sup> Z (5.84) plyne pro  $\tau = 0$  důležitý vztah

$$R_{xx}(0) = \int_{-\infty}^{\infty} S_{xx}(f) df. \quad (5.86)$$

Protože výraz na pravé straně (5.86) je celkový výkon signálu (viz str. 95), je hodnota autokorelační funkce v počátku rovna celkovému výkonu signálu (na odporu  $1\Omega$ ).

Příklady autokorelačních funkcí a odpovídajících (jednostranných) výkonových spektrálních hustot uvádí obr. 5.6.

Obr. 5.6 a ukazuje autokorelační funkci a výkonovou spektrální hustotu důležitého signálu zvaného *bílý šum*. Jde o signál, který ve svém spektru zastoupeny stejně složky všech frekvencí. Jeho název je odvozen z analogie s bílým světlem, které obsahuje složky všech barev světelného spektra. Viz též str. 99.

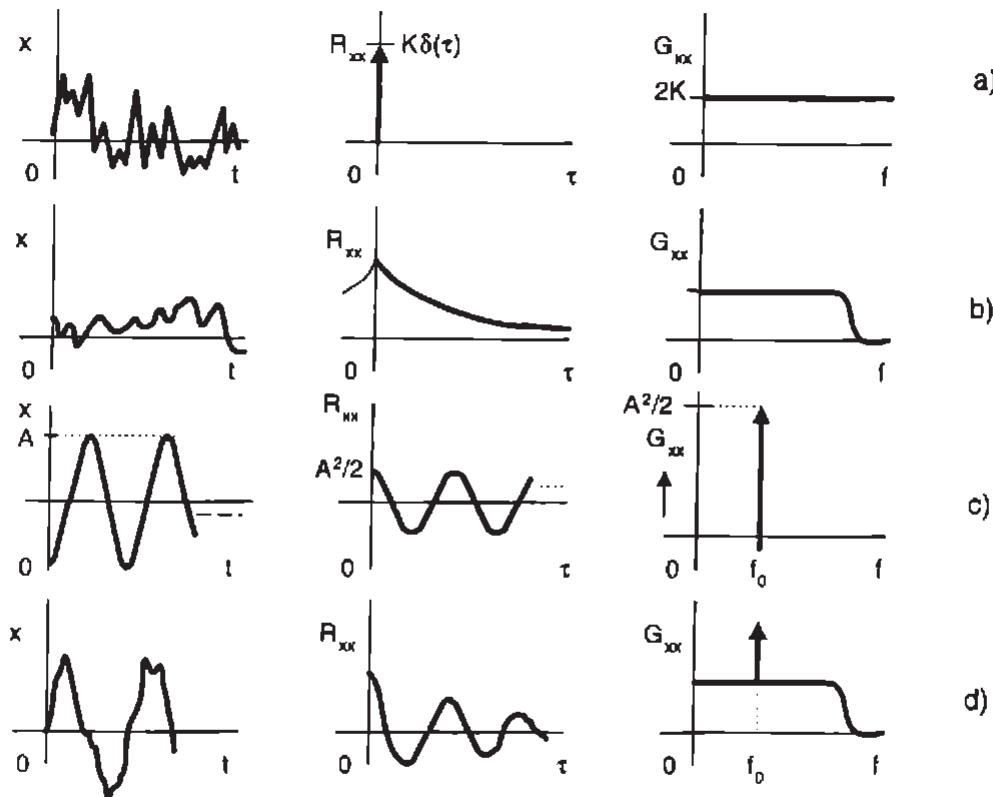
Z obr. 5.6 c je zřejmé, že autokorelační funkce nenese informaci o fázi (pro libovolný fázový posuv harmonického signálu je jeho autokorelační funkce dána výrazem  $(A^2/2) \cdot \cos \omega t$ ).

Pro signály časově omezené vyjde hodnota  $R_{xx}(\tau)$  podle (5.76) nulová. Tyto signály mají konečnou energii, jsou to tzv. *energetické signály*. Pro ně se autokorelační funkce definuje modifikovaným vztahem

$$\Phi_{xx}(\tau) = \lim_{T_M \rightarrow \infty} \int_0^{T_M} x(t)x(t + \tau) dt. \quad (5.87)$$

Signály časově neomezené, jejichž energie je nekonečná, se popisují výkonem a nazývají se *výkonové signály*. Časově neomezený signál je ovšem abstrakcí, v praxi pracujeme vždy s konečnými signály.

<sup>4</sup>Wienerovy-Chinčinovy vztahy lze psát také pro přímou Fourierovu transformaci (viz (5.117) a (5.118)).



Obrázek 5.6 Příklady realizací, autokorelačních funkcí a jednostranných výkonových spektrálních hustot: a) bílý šum, b) bílý šum po průchodu dolní propustí RC, c) harmonický signál s frekvencí  $f_0$ , d) harmonický signál s frekvencí  $f_0$  a s přidaným šumem

### 5.2.3 Užití korelačních funkcí

Dnes je na trhu řada aplikačních programů umožňujících mimo jiné výpočty korelačních funkcí. Příklady použití těchto funkcí uvádíme v této kapitole.

#### Užití vzájemné korelační funkce pro nalezení zpoždění mezi dvěma podobnými náhodnými signály

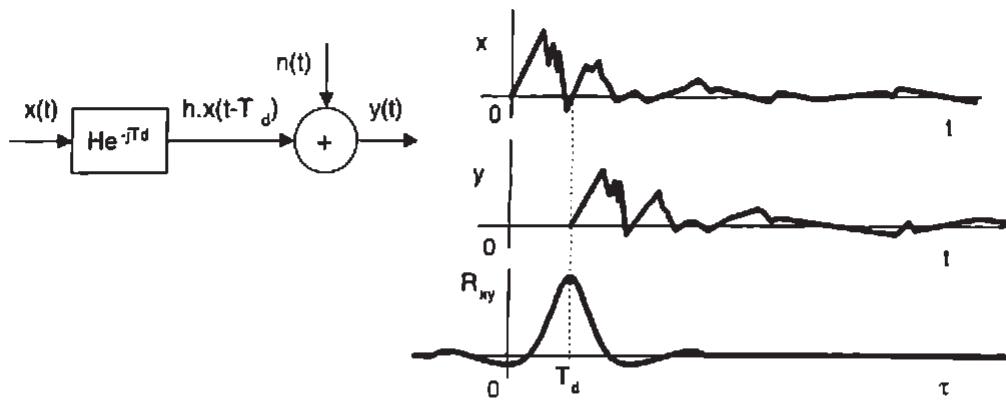
Předpokládejme, že signál  $x(t)$  se šíří rychlostí  $v$  v lineárním nedisperzním prostředí (pro které platí  $v \neq v(f)$ ) a výstupní signál  $y(t)$  sledujeme ve vzdálenosti  $d$  od zdroje signálu. Přitom se k signálu přidává aditivní šum  $n(t)$  nekorelovaný se signálem. Amplitudový frekvenční přenos prostředí  $|H(f)|$  je konstantní,  $H(f) = H \cdot e^{-jT_d \cdot 2\pi f}$ ,  $T_d = d/v$ . Platí

$$R_{xy}(\tau) = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_0^{T_M} x(t)[H \cdot x(t - \frac{d}{v} + \tau) + n(t + \tau)] dt = H R_{xx}(\tau - \frac{d}{v}). \quad (5.88)$$

Situaci demonstruje obr. 5.7.

Vzájemná korelační funkce je zde tedy úměrná autokorelační funkci posunuté o  $d/v$ , čili pro  $T_d = d/v$  má  $R_{xy}(\tau)$  maximum. Užití (5.88) je dvojí:

- Známe-li rychlosť  $v$  šíření signálu, můžeme určit vzdálenost  $d$ . Tento případ se využívá u radaru, sonaru nebo při ultrazvukovém vyšetřování v medicíně.
- Známe-li vzdálenost  $d$  (vzdálenost dvou snímačů signálu), můžeme určit rychlosť šíření  $v$ . Toho se využívá při měření rychlosti pohybu kapalin, sypkých hmot, plechů apod.



Obrázek 5.7 Užití vzájemné korelační funkce pro nalezení časového zpoždění

Signály z obou snímačů jsou náhodné, ale časově posunuté a případně částečně odlišné v důsledku turbulence proudění kapaliny apod.

### Zjištění periodicity signálu

Jde-li o periodický signál podobný náhodnému a je-li perioda relativně dlouhá, není periodicita ze záznamu zřejmá. Autokorelační funkce takového signálu je periodická se stejnou periodou a má poměrně výrazná maxima pro  $\tau = 0$  a  $\tau = k \cdot T_s$ , kde  $k$  je přirozené číslo a  $T_s$  je perioda signálu.

### Detecte periodického signálu v šumu, korelační filtrace

K tomuto účelu lze využít měření autokorelační nebo vzájemné korelační funkce. Předpokládáme, že na vstup měřicího zařízení působí signál  $x(t)$ , pro který platí

$$x(t) = s(t) + n(t), \quad (5.89)$$

kde  $s(t)$  je (užitečný) periodický signál a  $n(t)$  je ergodický šum, o kterém pro jednoduchost předpokládáme, že je bílý (viz obr. 5.6a a str. 99), že má nulovou střední hodnotu a normální rozložení amplitud. Předpokládáme také (v praxi splněno) nekorelovanost signálu se šumem.

Neznáme-li frekvenci předpokládaného signálu  $s(t)$ , měříme autokorelační funkci. Pokud frekvenci případně i tvar signálu známe, měříme vzájemnou korelační funkci. Situaci ukazuje obr. 5.8.

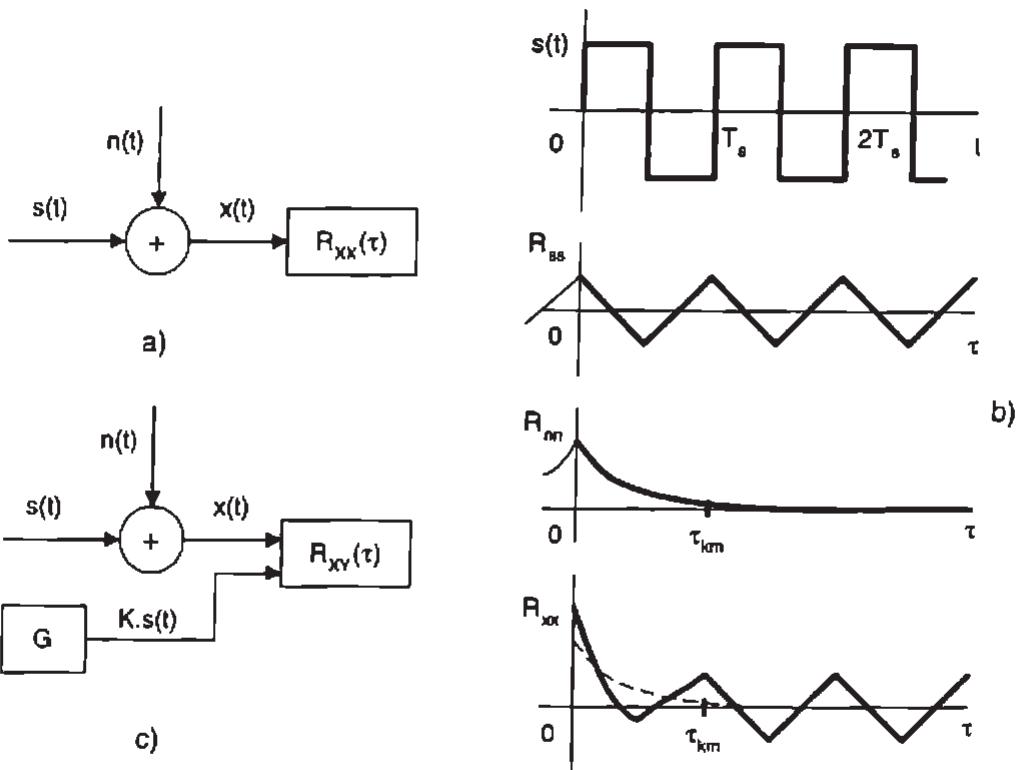
### Autokorelační filtrace

Pro odhad autokorelační funkce změřený korelátorem platí

$$\begin{aligned} \hat{R}(\tau) &= \frac{1}{T_M} \int_{T_M} [s(t) + n(t)][s(t + \tau) + n(t + \tau)] dt = \\ &= \hat{R}_{ss}(\tau) + \hat{R}_{sn}(\tau) + \hat{R}_{ns}(\tau) + \hat{R}_{nn}(\tau). \end{aligned} \quad (5.90)$$

Z předpokladu nekorelovanosti signálu a šumu a nulovosti střední hodnoty šumu plyne  $\hat{R}_{sn}(\tau) \rightarrow 0$ ,  $\hat{R}_{ns}(\tau) \rightarrow 0$ . Z předpokladu stacionárnosti šumu a nulovosti jeho střední hodnoty plyne

$$\lim_{\tau \rightarrow \infty} \hat{R}_{nn}(\tau) = 0. \quad (5.91)$$



Obrázek 5.8 Detekce periodického signálu v šumu pomocí měření autokorelační funkce (a, b) a vzájemné korelační funkce (c)

V praxi je vztah (5.91) splněn přibližně pro  $\tau > \tau_{km}$ , kde  $\tau_{km}$  je maximální interval korelace. Při měření  $R_{xx}(\tau)$  pro čas delší než je  $\tau_{km}$  tedy dostaneme vztah

$$\hat{R}_{xx}(\tau) \doteq \hat{R}_{ss}(\tau), \quad \tau > \tau_{km}. \quad (5.92)$$

Protože víme, že autokorelační funkce periodického signálu je také periodická a má stejnou periodu  $T_s$  jako signál, je v signálu  $x(t)$  užitečný periodický signál  $s(t)$  tehdy, je-li změřená autokorelační funkce periodická pro dostatečně velká posunutí  $\tau$ .

Tento metodou tedy zjistíme přítomnost periodického signálu v šumu a frekvenci tohoto signálu, ale nezjistíme tvar tohoto signálu. Frekvenci užitečného signálu nemusíme předem znát.

Tato metoda umožňuje detektovat periodický signál  $s(t)$  přibližně dvacetpětkrát slabší, než je šum. Z hodnoty  $R_{xx}(0)$  (absolutního maxima autokorelační funkce) a z výšky lokálních maxim periodického průběhu  $R_{xx}(nT_s)$  (pro  $\tau > \tau_{km}$ ) můžeme určit i poměr šumu k signálu, protože platí:

$$R_{xx}(0) = P_s + P_n, \quad R_{xx}(nT_s) = P_s, \quad (nT_s > \tau_{km}),$$

$$\frac{N}{S} = 10 \log \frac{P_n}{P_s} = 10 \log \frac{R_{xx}(0) - R_{xx}(nT_s)}{R_{xx}(nT_s)}. \quad (5.93)$$

Je tedy relativní rozdíl absolutního maxima a periodických lokálních maxim  $R_{xx}(\tau)$  měřítkem poměru šumu k signálu. Úroveň signálu rozeznatelného v šumu je dána dobou měření a rozlišovací schopnosti korelátoru. Rozlišovací schopnost korelátoru závisí na jeho dynamickém rozsahu. Měřicí rozsah musíme totiž volit tak, aby nedošlo k zahlcení (přetížení) vstupu korelátoru.

### Filtrace pomocí vzájemné korelační funkce

Tuto metodu použijeme, známe-li frekvenci signálu  $f_s$ , jehož přítomnost předpokládáme v šumu, případně i jeho tvar. V tomto případě použijeme na vstupu korelatoru (např. na přijímací straně telemetrického systému) generátor G generující průběh  $y(t) = K \cdot s(t)$ . Změříme-li vzájemnou korelační funkci signálů  $x(t) = s(t) + n(t)$  a  $y(t)$ , dostaneme

$$\hat{R}_{xy}(\tau) = \frac{1}{T_M} \int_{T_M} [s(t) + n(t)] \cdot K \cdot s(t + \tau) dt = K \cdot \hat{R}_{ss}(\tau) + K \cdot \hat{R}_{ns}(\tau). \quad (5.94)$$

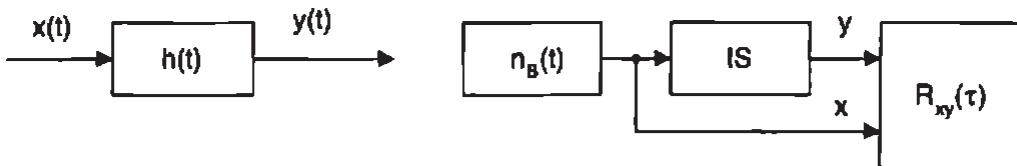
Protože předpokládáme nekorelovanost signálu  $s(t)$  se šumem, je druhý člen prakticky nulový. Platí tedy

$$\hat{R}_{xy}(\tau) \doteq K \cdot \hat{R}_{ss}(\tau). \quad (5.95)$$

Změřením vzájemné korelační funkce  $R_{xy}(\tau)$  tedy opět odhalí přítomnost periodického signálu utopeného v šumu, přičemž měření je  $K$ -krát citlivější než autokorelační filtrace. Měření je kratší než autokorelační filtrace, protože není nutno čekat na dostatečný pokles funkce  $R_{nn}(\tau)$ . Při příliš krátkém měření se ale může projevit statistická chyba - nenulovost odhadu  $\hat{R}_{ns}(\tau)$ . Výhodou proti autokorelační filtrace je i to, že zde nepúsobí rušivě harmonické signály o frekvencích  $f_r \neq f_s$ . Tato metoda se totiž chová jako filtrace úzkopásmovou propustí se střední frekvencí rovnou frekvenci signálu  $f_s$ .

### Identifikace soustav (nalezení impulsní odezvy)

Pustíme-li na vstup neznámého časově invariantního lineárního systému bílý šum  $n_B(t)$ , je vzájemná korelační funkce vstupu a výstupu tohoto systému  $R_{xy}(\tau)$  současně impulsní odezvou systému  $h(\tau)$  (viz obr. 5.9). To plyně z následujících vztahů.



Obrázek 5.9 Identifikace systému pomocí měření vzájemné korelační funkce ( $n_B(t)$  je bílý šum, IS je identifikovaný systém)

$$y(t) = \int_{-\infty}^{\infty} h(u) x(t-u) du \quad (5.96)$$

$$R_{xy}(\tau) = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_0^{T_M} x(t) \int_{-\infty}^{\infty} h(u) x(t+\tau-u) du dt = \quad (5.97)$$

$$= \int_{-\infty}^{\infty} h(u) R_{xx}(\tau-u) du = h(\tau) * R_{xx}(\tau). \quad (5.98)$$

Je tedy vzájemná korelační funkce konvolucí impulsní odezvy s autokorelační funkcí vstupního signálu.

Je-li vstupem bílý šum o jednotkové výkonové spektrální hustotě  $S_{xx}(f)$  (viz str. 95), je jeho korelační funkcí Diracův impuls v počátku ( $R_{xx}(\tau) = \delta(\tau)$ ). Platí tedy

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(u) \delta(\tau-u) du = h(\tau), \quad (5.99)$$

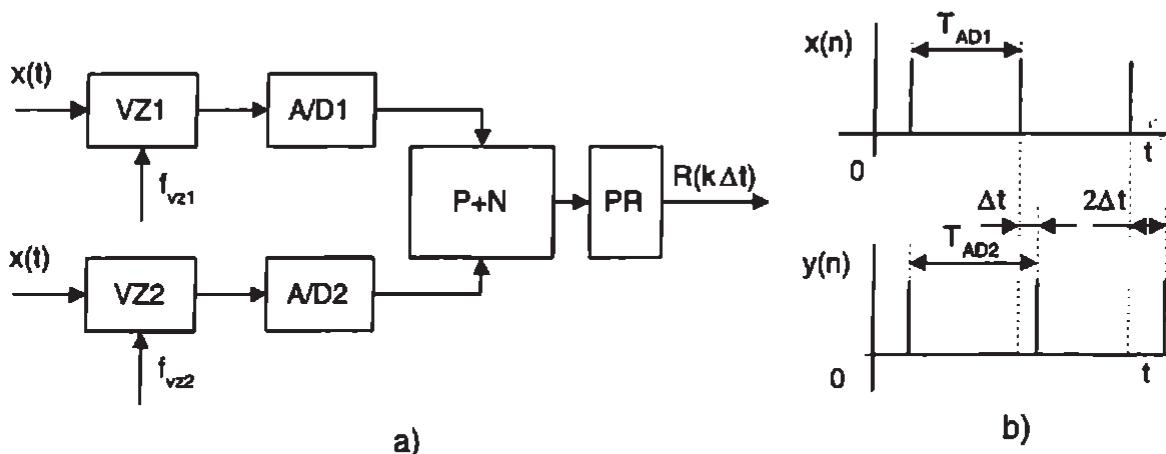
viz (2.18).

### 5.2.4 Měření korelačních funkcí

Korelační funkce můžeme měřit pomocí speciálních přístrojů - korelátorů, nebo vzorkováním průběhu a výpočtem na počítači. Druhý způsob je využíván i v mnohaúčelových *analyzátorech signálu*. Protože měření probíhá konečnou dobu  $T_M$ , zjistíme odhad korelační funkce. V praxi se předpokládá ergodičnost procesu a zpracovává se jedna realizace.

*Analogové korelátory* se dnes používají výjimečně. Měření probíhá podle (5.79). Rychlejší měření umožňují *reléový korelátor* a *znaménkový korelátor*, které pracují s jedním nebo oběma průběhy nahrazenými znaménkovou funkcí ( $\text{sign}[x(t)]$  resp.  $\text{sign}[y(t+\tau)]$ ). Více o analogových korelátorech viz [84].

*Číslicové korelátory* zpracovávají velké počty vzorků (např. až  $0,5 \cdot 10^6$ ), jsou přesnější než analogové a mohou měřit i pro velká zpoždění. Jejich princip včetně ukázky získání rozlišení podstatně převyšujícího vzorkovací interval pomocí principu nonia ukazuje obr. 5.10 (dle [84]). Jde o sekvenční vzorkování v ekvivalentním čase známé z teorie číslicových osciloskopů [23]. Řada číslicových korelátorů ale nonius nepoužívá a zpoždění v argumentu korelačních funkcí je dáno násobkem vzorkovacího intervalu.



Obrázek 5.10 Princip číslicového korelátoru (a) a nonia (b);  $VZ$  je vzorkovač,  $P + N$  je paměť a násobička,  $PR$  je průměrovací

*Maximální zpoždění*  $\tau_{MAX}$  při výpočtu korelačních funkcí se obvykle volí podstatně menší než doba měření  $T_M$ , obvykle  $\tau_{MAX} = T_M/10$ . Pracuje-li se se vzorkovanými signály, je zpoždění obvykle  $\tau = \tau \cdot T$ ,  $\tau = 0, 1, \dots, N/10$ . Má-li měření zachytit i malé výkyvy korelační funkce, doporučuje se volit krok zpoždění  $\Delta t \approx 1/(10 \cdot f_{MAX})$ , kde  $f_{MAX}$  je frekvence nejvyšší nezanedbatelné složky ve spektru signálu. Takové vzorkování je pětkrát rychlejší než vzorkování minimální frekvencí odpovídající vzorkovací větě. Dalším vodítkem pro volbu maximálního zpoždění může být hodnota odvozená z maximálního intervalu korelace  $\tau_{MAX} \approx (1 \div 1,5) \cdot \tau_{km}$ .

### Výpočet korelačních funkcí na počítači a v analyzátorech signálu

S rozšířením nabídky zásuvných karet do počítačů s A/D převodníky a rostoucím počtem analyzátorů signálu se stále více používá výpočet korelačních funkcí ze vzorkovaných realizací. Výpočet se provádí buď přímo z těchto vzorků, nebo nepřímo pomocí IFFT výkonové spektrální hustoty.

### Přímý výpočet korelačních funkcí ze vzorků realizace

Pracuje se s centrovanými posloupnostmi  $x(nT)$  a  $y(nT)$ , které získáme odečtením střední hodnoty od každého vzorku. Pak je  $R(\tau) \equiv K(\tau)$ . Doba měření je  $T_M = N \cdot T$ , máme k dispozici  $N$  vzorků. Korelační funkce počítáme pro zpoždění

$$\tau = r \cdot T, \quad r = 1, 2, \dots, m, \quad \text{kde } m = r_{\max} \leq 0,1N. \quad (5.100)$$

Definici (5.78) s vypuštěnou limitou nelze použít, protože nemáme k dispozici hodnoty pro  $n > N$ . Modifikujeme-li meze sumace, dostaneme pro kladná  $r$

$$\hat{R}_{xx}(rT) = \frac{1}{N} \sum_{n=1}^{N-r} x(nT) \cdot x(nT + rT). \quad (5.101)$$

Tento odhad je ale vychýlený a jeho střední hodnota dává autokorelační funkci upravenou trojúhelníkovým (Bartlettovým) oknem [45], pro které zde platí

$$w_T(r) = \frac{N-r}{N}, \quad r \leq N-1. \quad (5.102)$$

Konzistentní odhad dostaneme ze vztahu

$$\hat{R}_{xx}(rT) = \frac{1}{N-r} \sum_{n=1}^{N-r} x(nT) \cdot x(nT + rT), \quad r \ll N, \quad (5.103)$$

$$\hat{R}_{xx}(-rT) = \hat{R}_{xx}(rT). \quad (5.104)$$

Pro záporná zpoždění se tedy využije sudosti  $\hat{R}_{xx}(rT)$  pro reálná data. Pro  $r \leq N/10$  dává odhad (5.103) nepatrň větší disperzi než (5.101). Pro  $r \rightarrow N$  rozptyl odhadu (5.103) prudce roste, zatímco rozptyl (5.101) se blíží nule. Pro tato zpoždění ale  $\hat{R}_{xx}(rT)$  většinou nepočítáme<sup>5</sup>.

Vzájemné korelační funkce se počítají obdobně. Využívá se vztahu (5.72) a počítají se odhadы  $\hat{R}_{xy}(rT)$  a  $\hat{R}_{yx}(rT)$  pro kladná  $r$ . Platí

$$\hat{R}_{xy}(rT) = \frac{1}{N-r} \sum_{n=1}^{N-r} x(nT) \cdot y(nT + rT), \quad (5.105)$$

$$\hat{R}_{xy}(-rT) = \hat{R}_{yx}(rT). \quad (5.106)$$

### Nepřímý výpočet korelačních funkcí ze vzorků realizace s využitím FFT

Výraz (5.101) je podobný výrazu pro diskrétní konvoluci (liší se pouze znaménkem u  $nT$ ) a tedy by i pro tento výpočet bylo výhodné použít algoritmu pro rychlou konvoluci s využitím FFT. Protože ovšem DFT implicitně předpokládá periodičnost předmětu i obrazu, dostaneme výpočtem tzv. cyklickou korelační funkci  $R^c$ . Protože při výpočtu (aperiodické) konvoluce pomocí konvoluce kruhové musíme před výpočtem doplnit signály nulami minimálně na délku  $N_1 + N_2 - 1$ , je nutno při zpracování posloupnosti délky  $N$  pro výpočet  $m$  hodnot autokorelační funkce doplnit posloupnost nulami nejméně na délku  $L = N + m - 1$ . Protože se ale v praxi pro výpočet DFT používá většinou základních algoritmů FFT a protože výchozí délka realizace má obvykle  $N = 2^k$  vzorků, kde  $k$  je přirozené, doplňuje se před výpočtem realizace nulami na

<sup>5</sup>MATLAB ale počítá příkazem `xcorr` odhady korelačních funkcí pro všechn  $N$  hodnot, užitím parametru "unbiased" podle (5.103) a užitím parametru "biased" dle (5.101).

dvojnásobnou délku. Základem nepřímého výpočtu korelačních funkcí je Wienerův-Chinčinův vztah (str. 88) v diskrétní formě. V něm je odhad výkonové spektrální hustoty reprezentován periodogramem (viz str. 95). Pro autokorelační funkci pak platí

$$\hat{R}_{xx}(rT) = \text{IFFT} \{ [X(k\Delta f)]^* \cdot [X(k\Delta f)] \}. \quad (5.107)$$

Odhad korelační funkce se počítá následujícím postupem:

1.  $N$  hodnot realizace  $x(nT)$  doplníme na konci  $N$  nulami a vypočteme spektrum  $X(f_k)$ ; s uvážením vzorkovacího intervalu  $T$  platí

$$\frac{1}{T} X(f_k) = \sum_{n=0}^{2N-1} x(nT) e^{-j \frac{2\pi}{2N} nk}, \quad (5.108)$$

kde  $f_k = k \frac{f_w}{2N} = k \frac{1}{2NT} = k\Delta f$ .

2. Vypočteme tzv. nevyhlazený odhad jednostranné výkonové hustoty  $\hat{G}(f_k)$  (počítané jen pro nezáporné frekvence, viz str. 95 a dále) podle vztahu

$$\hat{G}_{xx}(f_k) = 2 \cdot \frac{1}{2NT} |X(f_k)|^2 = 2 \cdot \frac{T}{2N} \left| \frac{1}{T} X(f_k) \right|^2, \quad (5.109)$$

kde poslední úpravu jsme provedli proto, abychom mohli dosadit hodnoty získané přímo z FFT (viz (5.108)).

3. Zpětnou Fourierovou transformací  $\hat{G}_{xx}(f_k)$  dostaneme cyklickou autokorelační funkci  $R_{xx}^c$ , pro kterou platí

$$R_{xx}^c(rT) = \frac{N-r}{N} R_{xx}(rT) + \frac{r}{N} R_{xx}[-(rT - (2N-1)T)], r = 0, 1, \dots, N-1. \quad (5.110)$$

4. Pro získání  $m$  hodnot odhadu autokorelační funkce  $R_{xx}(rT)$ , kde  $m \approx N/10$  (výjimečně  $m \approx N/2$ ), upravíme prvních  $m$  hodnot  $R_{xx}^c(rT)$  podle vztahu

$$\hat{R}_{xx}(rT) = \frac{N}{N-r} R_{xx}^c(rT), \quad r = 0, 1, \dots, m. \quad (5.111)$$

Vzhledem k tomu, že autokorelační funkce reálných dat je sudá, je možno při výpočtu IFFT použít kosinovou transformaci. Průběh cyklické korelační funkce a oddělení jejich dvou složek doplněním realizace před výpočtem  $N$  nulami ukazuje obr. 5.11.

Výše uvedeným postupem lze počítat i vzájemné korelační funkce; obě realizace  $x(n)$  a  $y(n)$  opět doplníme na konci  $N$  nulami a místo s vlastní výkonovou spektrální hustotou  $G_{xx}$  pracujeme se vzájemnou výkonovou spektrální hustotou  $G_{xy}$  ([84, 62]).

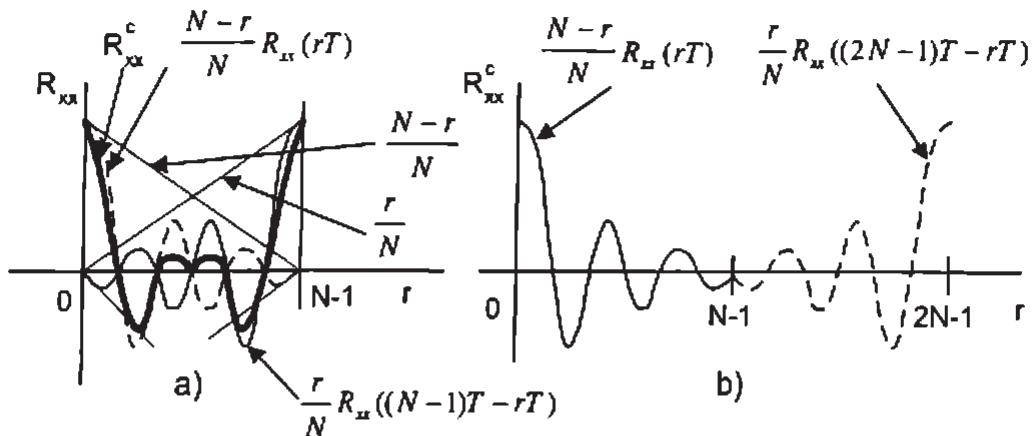
Tento postup výpočtu je rychlejší než výpočet přímo z realizace, pokud platí

$$m > 2 \cdot \log_2(2N) + 1. \quad (5.112)$$

V praxi je pro běžné rozsahy měření tento způsob rychlejší, protože  $m = 0,1N$  roste s  $N$  lineárně, zatímco podle (5.112) logaritmicky. Podmínka (5.112) je splněna přibližně pro  $N > 200$ .

### 5.3 Popis náhodných signálů ve frekvenční oblasti

Ve frekvenční oblasti se náhodné procesy popisují *výkonovými spektrálními hustotami*. V důsledku platnosti Wienerových-Chinčinových vztahů (5.84) a (5.85) je popis procesu pomocí výkonové spektrální hustoty ekvivalentní popisu pomocí korelačních funkcí. V praxi závisí na dostupné přístrojové technice a osobním vkusu, kterému popisu dáme přednost.



Obrázek 5.11 Dvě složky cyklické autokorelační funkce: a) při délce realizace  $N$ , b) po doplnění realizace nulami na délku  $2N$

### Definice a vlastnosti výkonových spektrálních hustot

*Výkonová spektrální hustota* charakterizuje rozdělení výkonu signálu v závislosti na frekvenci. Přibližně představuje výkon na odporu  $1\Omega$  v pásmu o šířce  $1Hz$ . Přesněji jde o limitu podílu výkonu v pásmu  $\Delta f$  k šířce tohoto pásmá, jde-li  $\Delta f$  k nule. V měřeních a jiných praktických aplikacích se používá *jednostranná výkonová spektrální hustota*  $G(f)$  nenulová pouze pro nezáporné frekvence. V teoretických úvahách využívajících Fourierovu transformaci se používá *dvoustranná výkonová spektrální hustota*  $S(f)$ .

Vlastní výkonová spektrální hustota  $S_{xx}(f)$  je definována buď jako střední hodnota tzv. periodogramu, nebo jako Fourierova transformace autokorelační funkce. Můžeme ji určit také analogovou filtrací a následným zpracováním realizace  $x(t)$ .

Pro dobu měření  $T_M$  lze pro každou realizaci vypočítat frekvenční spektrum

$$X_{TM}(f) = \int_0^{T_M} x(t) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} x_{TM}(t) e^{-j2\pi ft} . \quad (5.113)$$

*Periodogram* je výraz  $(1/T_M)|X_{TM}(f)|^2$  a výkonová spektrální hustota je střední hodnota v souboru periodogramů pro  $T_M \rightarrow \infty$ :

$$S'_{xx}(f) = \lim_{T_M \rightarrow \infty} E \left\{ \frac{1}{T_M} |X_{TM}(f)|^2 \right\} . \quad (5.114)$$

Jde tedy o reálnou nezápornou funkci. Pomocí jediné realizace lze výkonovou spektrální hustotu definovat vztahem

$$S_{xx}(f) = \lim_{\Delta f \rightarrow 0} \frac{1}{\Delta f} \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_{f-\Delta f/2}^{f+\Delta f/2} |X_{TM}(f)|^2 df . \quad (5.115)$$

Přitom je  $\Delta f \rightarrow 0$  a  $T_M \rightarrow \infty$  tak, že platí  $\Delta f \cdot T_M \rightarrow \infty$ . Odtud

$$S_{xx}(f) = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} |X_{TM}(f)|^2 . \quad (5.116)$$

Pro ergodické procesy platí  $S'_{xx}(f) = S_{xx}(f)$ .

Za definice výkonových spektrálních hustot (vlastní a vzájemné) lze považovat také *Wienerovy-Chinčinovy vztahy*<sup>6</sup>

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f\tau} d\tau, \quad (5.117)$$

$$S_{xy}(f) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-j2\pi f\tau} d\tau. \quad (5.118)$$

Dosazením  $f = 0$  do (5.117) dostaneme vztah obdobný (5.86):

$$S_{xx}(0) = \int_{-\infty}^{\infty} R_{xx}(\tau) d\tau. \quad (5.119)$$

Vzájemná výkonová spektrální hustota dvou signálů může být definována jako střední hodnota zobecněného periodogramu, v souboru nebo z jedné realizace

$$S'_{xy}(f) = \lim_{T_M \rightarrow \infty} E \left\{ \frac{1}{T_M} X_{TM}^*(f) Y_{TM}(f) \right\}, \quad (5.120)$$

$$S_{xy}(f) = \lim_{T_M \rightarrow \infty} \left\{ \frac{1}{T_M} X_{TM}^*(f) Y_{TM}(f) \right\}. \quad (5.121)$$

Je tedy zřejmé, že  $S'_{xy}(f)$  je komplexní funkce reálné proměnné  $f$ . Pro reálné signály  $x(t)$  platí

$$S_{xx}(-f) = S_{xx}(f) \geq 0, \quad (5.122)$$

$$S_{xy}(-f) = S_{xy}^*(f) = S_{yx}(f), \quad (5.123)$$

čili vlastní výkonová spektrální hustota reálného signálu je sudá a vzájemná výkonová spektrální hustota reálných signálů je konjugovaně sudá funkce  $f$ .

V praxi nemůžeme měřit pro záporné frekvence. Měřením dostaneme jednostranné výkonové spektrální hustoty  $G_{xx}(f)$  a  $G_{xy}(f)$ , pro které platí

$$G(f) = \begin{cases} 2S(f) & \text{pro } f \geq 0, \\ 0 & \text{pro } f < 0. \end{cases} \quad (5.124)$$

Třetím způsobem, jak můžeme získat odhad hodnoty výkonové spektrální hustoty, je propuštění signálu  $x(t)$  pásmovou propustí s šírkou pásma  $\Delta f$ , změření kvadrátu efektivní hodnoty napětí na výstupu této propusti a vydělení získané hodnoty veličinou  $\Delta f$ .

Označíme-li  $x(f, \Delta f, t)$  signál na výstupu ideální pásmové propusti o střední frekvenci  $f$  a šířce pásma  $\Delta f$ , platí

$$G_{xx}(f) = \lim_{\Delta f \rightarrow 0} \frac{1}{\Delta f} \left[ \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \int_{T_M} x^2(f, \Delta f, t) dt \right], \quad (5.125)$$

resp. pro odhad

$$\hat{G}_{xx}(f) = \frac{1}{(\Delta f) \cdot T_M} \int_{T_M} x^2(f, \Delta f, t) dt. \quad (5.126)$$

Výkonová spektrální hustota umožňuje nalézt výkon signálu v určitém frekvenčním pásmu. Výkon signálu v pásmu  $(f_1, f_2)$  vypočteme

$$P(f_1, f_2) = \int_{f_1}^{f_2} G_{xx}(f) df = \int_{-f_2}^{-f_1} S_{xx}(f) df + \int_{f_1}^{f_2} S_{xx}(f) df. \quad (5.127)$$

<sup>6</sup>Viz i (5.84) a (5.85).

Jednotkou výkonové spektrální hustoty je  $W \cdot Hz^{-1}$ , resp.  $V^2 \cdot Hz^{-1}$ . Protože se předpokládá signál na odporu  $1\Omega$ , vyjdou hodnoty v obou jednotkách stejně. V katalogích operačních zesilovačů se občas uvádějí průběhy  $\sqrt{G_{xx}}$  ve  $V/\sqrt{Hz}$ .

Velmi důležité v praktických aplikacích jsou vztahy mezi výkonovými spektrálními hustotami na vstupu (signál  $x$ ) a výstupu (signál  $y$ ) lineárního časově invariantního systému. Ze vztahu (5.98) a (5.124) dostaneme Fourierovou transformaci

$$S_{xy}(f) = H(f) \cdot S_{xx}(f), \quad (5.128)$$

$$G_{xy}(f) = H(f) \cdot G_{xx}(f). \quad (5.129)$$

Protože platí  $Y(f) = H(f)X(f)$  a tedy i  $|Y(f)|^2 = |H(f)|^2|X(f)|^2$ , dostaneme následujícím střední hodnoty druhého vztahu

$$S_{yy}(f) = |H(f)|^2 \cdot S_{xx}(f), \quad (5.130)$$

$$G_{yy}(f) = |H(f)|^2 \cdot G_{xx}(f). \quad (5.131)$$

*Výkonová spektrální hustota na výstupu lineárního časově invariantního systému je součin výkonové spektrální hustoty na vstupu tohoto systému a kvadrátu amplitudové frekvenční charakteristiky tohoto systému.*

Pro posloupnost vzorků ergodického náhodného signálu vzorkovaného v okamžicích  $nT$ ,  $n = -N, -(n-1), \dots, -1, 0, 1, \dots, N$  dostaneme DFT spektrum podle vztahu

$$X_{2N}(f_k) = X_{2N} \left( \frac{k}{(2N+1)T} \right) = T \cdot \sum_{n=-N}^{N} x(nT) e^{-j \frac{2\pi}{2N+1} nk}. \quad (5.132)$$

Zde jde o dvoustrannou časovou řadu délky  $2N + 1$ . Obdobný vztah můžeme napsat také pro jednostrannou řadu délky  $N$ , s jakou v praxi pracujeme. Odpovídající spektrum označíme  $X_N(f_k)$ . Obdobou vztahu (5.116) je pro posloupnost vzorků vztah

$$S_{xx}(f_k) = \lim_{N \rightarrow \infty} \frac{1}{(2N+1)T} |X_{2N}(f_k)|^2 = \lim_{N \rightarrow \infty} \frac{T}{2N+1} \left| \frac{X_{2N}(f_k)}{T} \right|^2. \quad (5.133)$$

Podobně pro  $S_{xy}(f_k)$  platí

$$S_{xy}(f_k) = \lim_{N \rightarrow \infty} \frac{1}{(2N+1)T} X_{2N}^*(f_k) Y_{2N}(f_k), \quad k = (-N, N), \text{ k-celé}. \quad (5.134)$$

Protože DFT (a FFT) předpokládá periodičnost průběhů  $x(n)$  i  $X(k)$  s periodou  $N$ , jsou spektrum dle (5.132) i výkonové spektrální hustoty vzorkovaných signálů periodické s periodou  $f_{vz} = 1/T$ , při použití argumentu  $\theta$  s periodou  $2\pi$ . Celkový výkon vzorkovaného signálu vypočteme pomocí této tzv. výkonové spektrální hustoty vzorkovaného signálu [45]  $S_{xx}(e^{j\theta})$  podle vztahu

$$P = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(e^{j\theta}) d\theta \quad (5.135)$$

a výkon ve frekvenčním pásmu  $(\theta_1, \theta_2)$  jako

$$P(\theta_1, \theta_2) = \frac{1}{2\pi} \int_{-\theta_2}^{-\theta_1} S_{xx}(e^{j\theta}) d\theta + \frac{1}{2\pi} \int_{\theta_1}^{\theta_2} S_{xx}(e^{j\theta}) d\theta. \quad (5.136)$$

Při výpočtech pomocí jednostranné vzorkované výkonové spektrální hustoty  $G_{xx}(e^{j\theta})$  je základní interval  $(0, \pi)$ , integrujeme jen pro  $\theta > 0$  a koeficient před integrály v (5.135) a (5.136) je  $1/\pi$ .

*Výkonová spektrální hustota bílého šumu* je konstantní. Zdroj ideálního analogového bílého šumu by měl mít nekonečný výkon ( $\int_{-\infty}^{\infty} S_{xx}(f) df \rightarrow \infty$ ). Takový zdroj samozřejmě neexistuje. V praxi stačí, má-li „bílý šum“ konstantní výkonovou spektrální hustotu ve frekvenčním pásmu  $\Delta f$  alespoň takovém, jaké zpracovává proměřovaný obvod. Takový signál se nazývá *frekvenčně omezený bílý šum* a platí pro něj

$$G_{xx}(f) = K = \frac{\sigma^2}{\Delta f}, \quad f \in \Delta f, \quad \Delta f = (0, f_h), \quad (5.137)$$

kde  $\sigma^2$  je celkový výkon šumu a  $\Delta f$  je frekvenční pásmo tohoto frekvenčně omezeného bílého šumu.

Průchodem bílého šumu s výkonovou spektrální hustotou  $G_{xx}(f) = K$  dolní propustí RC (pasivním integračním článkem) dostaneme na výstupu signál s výkonovou spektrální hustotou (viz (5.131))

$$G_{yy}(f) = |H(f)|^2 \cdot G_{xx}(f) = \frac{1}{1 + (2\pi f RC)^2} \cdot K. \quad (5.138)$$

Stejný výkon by měl signál na výstupu ideální dolní propusti s mezní frekvencí  $\Delta f_{s.ekv.}$  určenou podle vztahu

$$\Delta f_{s.ekv.} = \frac{1}{|H_{max}|^2} \int_0^{\infty} |H(f)|^2 df. \quad (5.139)$$

Hodnota  $\Delta f_{s.ekv.}$  je *ekvivalentní šířka pásma* (v tomto případě dolní propusti, ale obdobně se definuje i pro pásmovou propust). Jde o šířku pásma ideálního frekvenčně selektivního filtru, kterým nahrazujeme reálný filtr, který při bílém šumu na svém vstupu dává na výstupu stejný výkon jako nahrazovaný reálný filtr.

Tvrzení o fyzikální nerealizovatelnosti bílého šumu ale neplatí pro *uzorkovaný (diskrétní) bílý šum e(n)*. Jeho výkon se spočítá pomocí (5.135) a je konečný. Jeho střední hodnota je nulová, jeho vzorkovaná výkonová spektrální hustota je rovna jeho disperzi  $\sigma^2 = E\{e^2(n)\}$ . Jeho autokorelační funkce je nenulová jen v počátku, kde je rovna  $\sigma^2$ . Tento šum je fyzikálně realizovatelný [63].

Harmonický (sinusový) signál s frekvencí  $f_0$  a amplitudou  $A$  má průběh  $G_{xx}(f)$  tvořený Diracovým impulsem na frekvenci  $f_0$  s váhou  $A^2/2$  (obr. 5.6). Průběh  $S_{xx}(f)$  tvoří dvojice Diracových impulsů na frekvencích  $f_0$  a  $-f_0$  s váhami  $A^2/4$ .

### 5.3.1 Způsoby získání výkonových spektrálních hustot

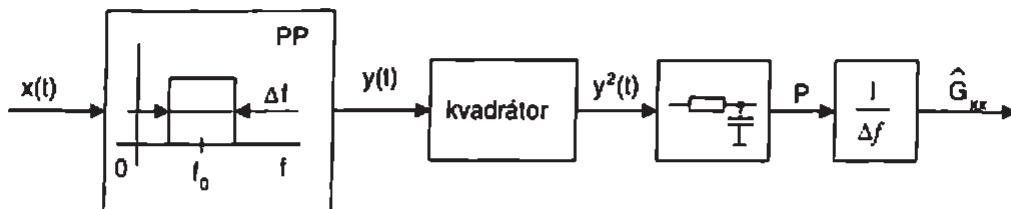
#### Měření spektrální hustoty analogovou filtrací

Ergodický náhodný signál  $x(t)$  se filtruje pásmovou propustí s šířkou pásma  $\Delta f$  a nastavitelnou střední frekvencí  $f_0$ . Změřením kvadrátu efektivní hodnoty výstupního napětí filtru a jeho vydelením hodnotou  $\Delta f$  dostaneme odhad  $\hat{G}_{xx}(f)$  (obr. 5.12).

Pro odhad  $\hat{G}_{xx}(f)$  platí ( $T_M$  je doba měření)

$$\hat{G}_{xx}(x) = \frac{1}{\Delta f_{s.ekv.}} \frac{1}{T_M} \int_{T_M} [x(t)]_{\Delta f_{s.ekv.}, f_0}^2 dt. \quad (5.140)$$

Měníme-li  $f_0$ , lze změřit postupně celý průběh  $\hat{G}_{xx}(f)$ . Současné měření několika hodnot umožňují vícekanálové analyzátory. Pásmová propust z obr. 5.12 může mít buď konstantní šířku



Obrázek 5.12 Blokové schéma analogového měření výkonové spektrální hustoty  $G_{xx}(f)$

pásma (využívá-li se heterodynmiho principu), nebo stálou poměrnou šířku pásma  $\Delta f_{s.e.kv.} = k f_0$  (když je filtr tvořen pásmovou zádrží tvořenou dvojitým T - článkem ve zpětné vazbě zesilovače).

### Výpočet výkonové spektrální hustoty metodou modifikovaných periodogramů

Dnes se pracuje většinou se vzorkovanými průběhy. Nejrozšířenější metodou výpočtu výkonové spektrální hustoty je *metoda modifikovaných periodogramů*, případně *metoda Welchova* [98]. Používá se v signálních analyzátorech i při zpracování signálů ze zásuvných karet s A/D převodníky na počítačích.

Předpokládáme centrovany ergodický náhodný proces, realizace má délku  $T_M = NT$ . Odhadem  $G_{xx}(f)$  je zde periodogram. Ten je ale odhadem nekonzistentním, proto se vyhlažuje zprůměrováním několika periodogramů kratší délky. Ty získáme rozdelením realizace (posloupnosti délky  $N$ ) na  $M$  stejně dlouhých segmentů, které se buď neprekryvají, nebo se mohou (zpravidla maximálně z 50%) překrývat. Každý ze segmentů před dalším zpracováním vynásobíme vhodným oknem. Nejčastěji se zde používá okno Hann (4.55). Násobení oknem je možno provést také výpočtem konvoluce spektra signálu a spektra okna ve frekvenční oblasti. To je u okna Hannova velmi jednoduché (viz str. 39).

*Postup nalezení výkonové spektrální hustoty metodou modifikovaných periodogramů* je následující:

1. Z posloupnosti vzorů  $x'(nT)$  vypočteme vzorky centrového signálu  $x_c(nT) = x'(nT) - \mu_x$ .
2. Rozdělíme posloupnost  $N$  centrových vzorků na  $M$  segmentů délky  $L$  (pokud jde o překrývající se segmenty, platí  $L \cdot M > N$ ).
3. Každý segment vynásobíme oknem délky  $L$ , např.  $w_{HN}$ , takže dál pracujeme s daty  $x(nT) = x_c(nT) \cdot w_{HN}(nT)$ .
4. Vypočteme spektrum každého segmentu  $X_i(nT)$  pomocí FFT

$$X_{iT}(k/LT) = T \cdot \sum_{n=0}^{L-1} x_i(nT) e^{-j \frac{2\pi}{L} n \cdot k}. \quad (5.141)$$

5. Hrubý odhad vzorkované výkonové spektrální hustoty  $i$ -tého segmentu dat  $\hat{G}_{ixx}(f_k)$  pro  $f_k = k/(LT)$  vypočteme podle vztahu

$$\hat{G}_{ixx} \left( \frac{k}{LT} \right) = \hat{G}_{xx}(f_k) = 2 \cdot \frac{1}{LT} |X_T(f_k)|^2 = 2 \cdot \frac{T}{L} \left| \frac{1}{T} X_{iT}(f_k) \right|^2. \quad (5.142)$$

Poslední úprava je provedena proto, že aplikováním FFT na posloupnost  $x(nT)$  dostaneme  $(1/T)(X_T(f_k))$ , viz (5.141).

6. Odhad  $\hat{G}_{xx}(f_k)$  je třeba vyhladit. *Vyhlazený odhad* dostaneme výpočtem průměru všech  $M$  segmentů:

$$\tilde{G}_{xx}(f_k) = \frac{1}{M} \sum_{i=0}^{M-1} \hat{G}_{ixx}(f_k). \quad (5.143)$$

Tímto výpočtem dosáhneme  $M$ -násobného zmenšení rozptylu odhadu ( $\sigma$  se zmenší  $(1/\sqrt{M})$ -krát).

7. Úpravou signálu časovým oknem se zmenší hodnoty vzorků realizace. Aby se zachovala původní efektivní hodnota a výkon signálu, čili abychom dostali nevychýlený odhad, je nutno násobit hodnoty vyhlazeného odhadu konstantou  $1/\sum w^2(nT)$ , kde  $w(nT)$  je použité okno.

### Výpočet výkonové spektrální hustoty zpětnou Fourierovou transformací kovarianční funkce

Tato metoda se používala zejména před objevem algoritmů FFT. Podle autorů se nazývá *metoda Blackmanova-Tukeyova*. Dnes se ovšem i v této metodě může využívat algoritmu FFT.

Vychází se z korelační funkce, která se získá buď výpočtem ze vzorků realizace (vztahy (5.103) a (5.105)) nebo změřením na číslicovém korelátoru. Protože  $R_{xx}(\tau)$  známe pro zpoždění 0 až  $\tau_M = m \cdot T$ , zpracováváme hodnoty upravené obdélníkovým oknem  $w_o(\tau)$ . Protože spektrum tohoto okna nabývá i záporných hodnot, mohou z inversní Fourierovy transformace součinu  $R_{xx}(\tau) \cdot w_o(\tau)$ , čili z konvoluce ve frekvenční oblasti

$$\hat{G}_{xx}(f) = W_o(f) * G_{xx}(f) \quad (5.144)$$

vyjít i záporné hodnoty nevyhlazeného odhadu  $\hat{G}_{xx}(f)$ , i když víme, že musí být  $G_{xx}(f) > 0$ . Proto je opět výhodnější použít jiná okna, která mají nižší postranní laloky, i když je jejich hlavní lalok širší. Zde se používají okna souměrná kolem počátku. Okno Hannovo a Hammingovo můžeme vyjádřit společným vztahem

$$w(\tau) = \begin{cases} \alpha_0 + 2\alpha_1 \cdot \cos\left(\frac{\pi\tau}{\tau_m}\right), & |\tau| < \tau_m \\ 0, & \text{jinde.} \end{cases} \quad (5.145)$$

Úpravu oknem opět můžeme provést ve frekvenční oblasti. S využitím linearity FT, vyjádření kosinu pomocí exponenciál, a užitím věty o posuvu v obrazu FT dostaneme pro okna ze vztahu (5.145)

$$w(\tau) \cdot R_{xx}(\tau) \xleftrightarrow{FT} W(f) * \hat{G}_{xx}(f), \quad (5.146)$$

$$(\alpha_0 + 2\alpha_1 \cos\frac{\pi\tau T}{mT}) \cdot R_{xx}(rT) \xleftrightarrow{FFT}$$

$$\xleftrightarrow{FFT} \alpha_0 \hat{G}_{xx}(f_k) + \alpha_1 [\hat{G}_{xx}(f_{k-1}) + \hat{G}_{xx}(f_{k+1})],$$

kde  $f_k = k/(mT)$ ,  $k = 0, 1, \dots, (m-1)$ .

Přitom  $\hat{G}_{xx}(f_k) = FFT \left\{ \hat{R}_{xx}(rT) \right\}$ . Protože funkce  $R_{xx}(rT)$  je sudá, je možno pro výpočet FFT použít kosinovou transformaci, která je rychlejší než klasické algoritmy FFT. Z uvedeného je zřejmé, že úprava nevyhlazeného odhadu oknem spočívá ve výpočtu tříčlenného váženého klouzavého průměru. Aplikace okna tedy znamená vyhlazení průběhu  $\hat{G}_{xx}$ .

*Postup výpočtu  $G_{xx}(f)$  metodou Blackman-Tukey* je tedy následující:

1. Centrování hodnot realizace (odečtení  $\mu_x$  od všech hodnot).
2. Výpočet odhadu  $\hat{R}_{xx}(rT)$  (vztah (5.103)).
3. Výpočet nevyhlazeného odhadu  $\hat{G}_{xx}(f_k)$  - FFT průběhu  $\hat{R}_{xx}(rT)$ .
4. Výpočet vyhlazeného odhadu  $\tilde{G}_{xx}(f_k)$ .

Nejdelší je výpočet podle bodu 2. Urychlit je ho možné nepřímým výpočtem s využitím FFT (viz str. 94).

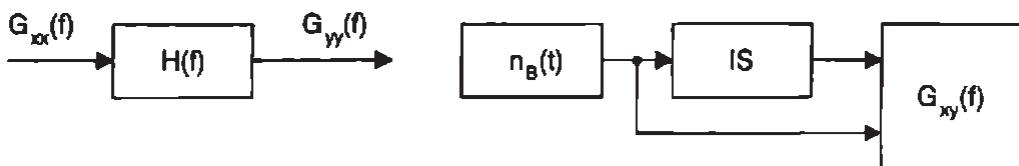
Oběma způsoby je možné počítat také odhady vzájemných výkonových spektrálních hustot. Ty jsou ale komplexními funkciemi, takže se musejí zjišťovat zvlášť reálné a imaginární části hodnot. Protože většina programů pro výpočet FFT využívá algoritmu pro  $N = 2^m$ ,  $m$  – přirozené, je nutno data před zpracováním na tuto délku buď zkrátit nebo doplnit nulami.

### 5.3.2 Použití výkonových spektrálních hustot

Jednou z aplikací je měření vzájemné výkonové spektrální hustoty pro zjištění frekvenční charakteristiky lineární soustavy. Jde o přímou obdobu nalezení impulsní odezvy soustavy podle obr. 5.9. Z (5.129) plyne

$$H(f) = \frac{G_{xy}(f)}{G_{xx}(f)}. \quad (5.148)$$

Je-li  $x(t)$  bílý šum, pro který platí  $G_{xx}(f) = K_1$ , je frekvenční odezva soustavy  $H(f)$  přímo úměrná  $G_{xy}(f)$ , neboť z (5.148) plyne pro tento případ  $H(f) = (1/K_1)G_{xy}(f)$ . Výhodou identifikace soustav dle obr. 5.13 je možnost měření za normálního provozu soustavy, kdy je bílý šum přidán jako aditivní k pracovnímu vstupnímu signálu soustavy a je měřena vzájemná výkonová spektrální hustota vstupního šumu a výstupního signálu soustavy.



Obrázek 5.13 Nalezení frekvenční charakteristiky lineární soustavy měřením  $G_{xy}(f)$ ;  $n_B(t)$  je bílý šum, IS je identifikovaná soustava

Dalším použitím výkonových spektrálních hustot je experimentální určení *koherenční funkce*. Koherenční funkce je definována vztahem

$$\gamma_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f) \cdot G_{yy}(f)}. \quad (5.149)$$

Z (5.149) je zřejmé, že by mělo platit  $\gamma_{xy}(f) = 1$ . Při měření tato funkce často vyjde  $\gamma_{xy}(f) < 1$ . Důvodem může být přítomnost vnějšího šumu v měřicím obvodu, nelineárnost soustavy, vychýlenost odhadů  $G_{xx}$ ,  $G_{xy}$  a  $G_{yy}$  nebo ovlivnění výstupu  $y(t)$  i jinými vstupními procesy než pouze  $x(t)$ . Pokud měříme v závislosti na frekvenci několik různých funkcí (např.  $|H(f)|$ ,  $G_{xy}(f)$  apod.) a současně měříme koherenční funkci, pak pro frekvence, kde je  $\gamma_{xy}(f) < 1$ , je věrohodnost ostatních měření snížena. Víc o koherenční funkci viz [2].

## 5.4 Měření signálů za přítomnosti šumů

V měřicí i řídící technice velmi často působí na vstup zařízení vedle signálu také šum. Šum je typickým stochastickým procesem a jako takový je charakterizován amplitudovým rozložením a

korelační funkcí nebo výkonovou spektrální hustotou. Typickým představitelem šumu je *teplotní šum odporu*, pro jehož efektivní hodnotu platí

$$U_{ef.} = \sqrt{4 k T R \Delta f}, \quad (5.150)$$

kde  $k$  je Boltzmannova konstanta ( $k = 1,38 \cdot 10^{-23} J \cdot K^{-1}$ ),  $T$  je absolutní teplota v  $K$ ,  $R$  je odpor a  $\Delta f$  je frekvenční pásmo (zpracovávané zařízením, na jehož výstupu šum zjišťujeme).

Šumy se objevují na výstupu všech elektronických obvodů. Kromě toho se k pracovním nebo měřeným signálům v praxi přidává vnější rušení elektromagnetickou indukcí nebo kapacitní vazbou. Pro zpracování signálů za přítomnosti šumů se používají *tři základní skupiny metod*:

1. *Korelační filtrace.*
2. *Průměrování signálu.*
3. *Frekvenční filtrace.*

Za zvláštní a velmi účinnou metodu korelační případně frekvenční filtrace lze považovat *synchronní detekci*.

Zpracování zašuměných signálů spočívá buď v *detekci užitečného signálu* (čili zjištění přítomnosti periodického užitečného signálu v šumu), důležité zejména v radiotechnice, nebo ve *vydělení užitečného signálu ze šumu*. V obou případech bývá užitečný signál podstatně slabší než šum, takže při zobrazení na osciloskopu není vůbec patrný. Dále popsanými metodami můžeme zjistit jeho přítomnost, případně i jeho průběh.

Kritériem posouzení odolnosti systému vůči poruchám případně posouzení účinnosti jednotlivých metod je minimální hodnota poměru signálu k šumu, kdy ještě systém funguje nebo metoda je účinná. *Poměr signálu k šumu* je definován vztahy

$$\frac{S}{N} = 20 \cdot \log \frac{U_{ef.s.}}{U_{ef.n.}} [\text{dB}] \quad (5.151)$$

$$\frac{S}{N} = 10 \cdot \log \frac{P_s}{P_n} [\text{dB}], \quad (5.152)$$

kde  $S$  označuje signál a  $N$  označuje šum a jeho parametry (z anglicky *noise*). Někdy se pracuje přímo s (bezrozměrným) poměrem  $P_s/P_n$ .

Metody pro zpracování signálu se šumem jsou tím účinnější, čím menší je hodnota  $S/N$ , pro kterou je metoda ještě účinná. Při rozboru metod předpokládáme v praxi nejběžnější případ, že se šum k signálu přičítá (*aditivní šum*)<sup>7</sup>, je stacionární a ergodický, nekorelovaný se signálem, má nulovou střední hodnotu a je bílý. Navíc pro jednoduchost předpokládáme, že jde o gaussovský šum.

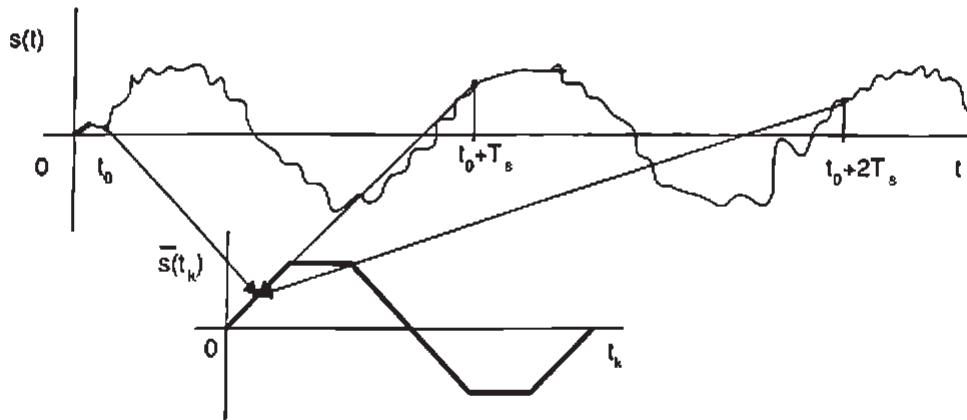
### 5.4.1 Korelační filtrace

Korelační filtrace umožňuje detektovat periodický signál v šumu. Spočívá v měření autokorelační funkce (pokud neznáme frekvenci periodického signálu předpokládaného v šumu) nebo v měření vzájemné korelační funkce signálu smíchaného se šumem s pomocným signálem stejné frekvence (případně i tvaru), jako má očekávaný signál. Popis této metod i jejich srovnání je uveden na str. 90 a 92.

<sup>7</sup>V praxi ale existují i případy, kdy rušení ovlivňuje přenosovou konstantu a působí tedy multiplikativně. To je např. případ vlivu atmosférických poruch na přenos radiových signálů.

### 5.4.2 Průměrování signálu

Pomocí průměrování signálu (vybíráni signálu ze šumu, angl. *signal enhancement*) zjistíme nejen přítomnost užitečného *periodického signálu*  $s(t)$  v měřeném signálu  $x(t) = s(t) + n(t)$ , ale i tvar signálu  $s(t)$ . Průměrování je součástí algoritmů zpracování signálu pro zvýšení rozlišení číslicových osciloskopů a přesných stejnosměrných laboratorních číslicových voltmetrů. Princip jednoduchého (sumačního) průměrování naznačuje obr. 5.14.



Obrázek 5.14 Příklad vzorkování periodického signálu pro metodu průměrování

Vzorkovací interval se volí podle vztahu  $T = T_s$ , kde  $T_s$  je perioda signálu. Pro ovládání vzorkovače je nejvhodnější použít vnější spouštění synchronizované se signálem  $s(t)$ . Není-li signál utopen v šumu, lze použít i vnitřní spouštění. Odebereme  $N$  vzorků signálu  $x(t)$  a z nich vypočteme aritmetický průměr:

$$\begin{aligned}\overline{x_N(t_0)} &= \frac{1}{N} \sum_{i=1}^N [s(t_0 + iT_s) + n(t_0 + iT_s)] = \\ &= \frac{1}{N} N s(t_0) + \frac{1}{N} \sum_{i=1}^N n(t_0 + iT_s) = s(t_0) + \hat{\mu}_N \doteq s(t_0),\end{aligned}\quad (5.153)$$

protože platí  $\lim_{N \rightarrow \infty} \hat{\mu}_N = 0$ .

Při změně hodnoty  $t_0$  v rozsahu celé periody dostaneme celý průběh signálu. Rekonstrukci signálu lze provádět také mnohakanálově a tedy podstatně rychleji, pokud budeme popsanou metodou určovat větší množství hodnot  $x_N(t_0)$  najednou. Například korelátor SEM NORMA takto zjišťuje 100 bodů s volitelnou hodnotou  $N$  v rozsahu  $2^7$  až  $2^{19}$ .

V praxi je  $N$  vždy konečné, a proto není průměrováním šum zcela eliminován. Jsou-li sousední vzorky šumu nekorelované, zmenší se směrodatná odchylka šumu z hodnoty  $\sigma_n$  průměrováním na  $\sigma_n / \sqrt{N}$ . Signál je průměrováním neovlivněn, takže se  $\sqrt{N}$ -krát zvýší i poměr S/N.

Nevýhodou popsaného způsobu průměrování je nutnost před výpočtem průměru čekat na odebrání všech vzorků. Zrychlení lze dosáhnout *rekurzivním průměrováním* podle vztahů

$$\overline{x_0(t_0)} = 0, \quad (5.154)$$

$$\overline{x_i(t_0)} = \frac{1}{i} x_i + \frac{i-1}{i} \overline{x_{i-1}(t_0)}, \quad i = 1, 2, \dots \quad (5.155)$$

## Průměrování klouzavým průměrem

K vyhlazení šumu lze použít také *klouzavý průměr*, zvláštní typ transversálního FIR filtru popsaný na str. 53. Zde se obvykle počítá průměr v okolí každého výsledného bodu z podstatně menšího počtu bodů než tvoří celková délka signálu. Je-li frekvence signálu postatně menší než je frekvence první nuly filtru (čili než  $f_{vz}/L$ , kde  $L$  je délka filtru), není signál filtrem prakticky zeslaben, ale širokopásmový šum je potlačen značně. To je zřejmé z průběhu frekvenční charakteristiky tohoto filtru (obr. 4.14). Působí-li společně s harmonickým užitečným signálem o frekvenci  $f_s$  periodické rušení s frekvencí  $f_r > f_s$ ,  $f_r \neq m \cdot f_s$ , lze je potlačit volbou délky filtru tak, aby  $f_r$  padla do některé nuly frekvenční charakteristiky filtru.

Na str. 54 je popsán klouzavý průměr s exponenciálním zapomínáním, který je diskrétní obdobou filtrace pasivním integračním článkem. Tento filtr se užívá v analyzátorech signálu a číslicových korelátorech v režimu zvaném *exponenciální průměrování*. Vzorky jsou započítávány s tím menší váhou, čím je vzorek vzdálenější od právě odebíraného vzorku, proto se tento režim nazývá také *průměrování se zapomínáním*. Od *lineárního průměrování* se liší tím, že signál filtruje trvale (filtrace neskončí po odebrání  $N$  vzorků).

### 5.4.3 Frekvenční filtrace

Možnost zvýšení poměru  $S/N$  frekvenční filtrací závisí na vzájemné poloze frekvenčních spekter užitečného signálu  $s(t)$  a šumu (rušení)  $n(t)$ . *Princip frekvenční filtrace* spočívá v omezení frekvenčního pásma měřeného signálu  $x(t) = s(t) + n(t)$  tak, aby nebylo omezeno (bylo minimálně omezeno) frekvenční spektrum užitečného signálu  $s(t)$  a bylo maximálně omezeno frekvenční pásmo s nenulovou spektrální hustotou šumu. Tím dojde ke zvýšení poměru  $S/N$ .

Nejúčinnější je tato metoda v případě, kdy se spektra signálu a šumu nepřekrývají. Pak se signál propustí a šum potlačí vhodným frekvenčně selektivním filtrem (obr. 4.12). Při částečném překrytí spekter signálu a šumu je filtrací ovlivněn i tvar signálu.

Předpokládáme-li bílý šum, bude výkon tohoto šumu lineárně snižován při lineárním snižování šířky propustného pásma filtru (dolní propusti nebo pásimové propusti). Měříme-li zařízením (např. osciloskopem) s frekvenčním pásmem  $\Delta f_m = (f_{dm}, f_{hm})$ , ve kterém leží celé spektrum signálu, není výkon signálu zařízením omezen. Výkon změřeného (bílého) šumu je roven  $K \cdot (f_{hm} - f_{dm})$ , kde  $K = G_{xx}$  je jednostranná výkonová spektrální hustota bílého šumu. Předřadíme-li měřicímu zařízení filtr s propustným pásmem  $\Delta f_F = (f_{dF}, f_{hF})$  tak, aby neovlivnil výkon signálu, zvětšíme poměr  $S/N$  měřeného signálu  $x(t)$  o hodnotu

$$\left( \frac{S}{N} \right)' - \frac{S}{N} = 10 \cdot \frac{f_{nm} - f_{dm}}{f_{hF} - f_{dF}} \quad (\text{dB}). \quad (5.156)$$

Tato metoda je mimořádně vhodná pro harmonický (sinusový) signál, případně signál periodický. Snižováním šířky pásmové propusti při měření výkonové spektrální hustoty lze teoreticky libovolně zvýšit poměr signálu k šumu<sup>8</sup>. To ale předpokládá, že (jednostranné) spektrum harmonického signálu je tvořeno jedinou čarou, což je splněno jen pro časově neomezený signál. Při časově omezeném měření vzorkovaného signálu se obecně projeví vlivem konvoluce (ideálního) spektra sinusovky se spektrem obdélníkového okna a vzorkování výsledku ve frekvenční oblasti rozmažání spektra (*leakage*, viz str. 26). Pro získání spektra, které prakticky celé leží v propustném pásmu  $\Delta f_{pp}$  použité dolní propusti, je proto nutno měřit dostatečně

<sup>8</sup>Pokud k filtraci použijeme dolní propust s propustným pásmem  $(0, f_{sig})$ , projeví se šum v propustném pásmu filtru typicky kolísáním amplitudy sinusového signálu.

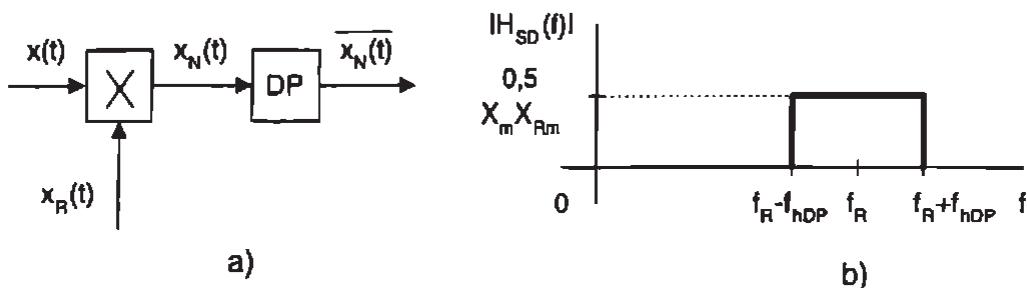
dlouho a k dispozici musí být výpočetní zařízení s odpovídající kapacitou. Pokud toto není splněno, je nutné použít speciálních metod detekce slabého signálu v šumu, jako Pronyho metody, Pisarenkova harmonického rozkladu, nebo použití modelů AR nebo ARMA (viz např. [46]).

#### 5.4.4 Synchronní (koherentní) detekce

*Synchronní detekce* využívá obvod zvaný synchronní modulátor (viz obr. 5.15 a) nebo funkci tohoto modulátoru simuluje číslicově. Jde o mimořádně účinnou metodu detekce harmonického nebo jiného periodického signálu utopeného v šumu, protože umožní zjistit přítomnost tohoto signálu v šumu až pro poměr  $S/N = -100 \text{ dB}$ . Princip spočívá v násobení signálu  $x(t)$  referenčním harmonickým signálem  $x_R(t)$  o frekvenci  $f_R$  a následné filtraci výsledného signálu dolní propustí. Jde tedy vlastně o amplitudovou modulaci signálu  $x(t)$  signálem  $x_R(t)$ ,<sup>9</sup> která přenese frekvenční pásmo signálu  $\Delta f_s$  do okolí frekvence  $f_R$ . Tuto frekvenci můžeme zvolit v pásmu, kde je slabé vnější rušení nebo aditivní šum  $n(t)$  tvořící součást signálu  $x(t)$ .

Mezní frekvence dolní propusti se volí  $f_{hDP} < 0,5 \cdot f_R$ . Pro sinusový vstupní signál  $x(t) = X_m \cdot \sin(2\pi f_s t)$  a referenční signál  $X_R(t) = X_{Rm} \cdot \sin(2\pi f_R t + \varphi)$  platí

$$x_v(t) = \frac{1}{2} X_m \cdot X_{Rm} \{ \cos[(2\pi f_R - 2\pi f_s)t - \varphi] - \cos[(2\pi f_R + 2\pi f_s)t + \varphi] \}. \quad (5.157)$$



Obrázek 5.15 Synchronní modulátor (a) a jeho amplitudová frekvenční charakteristika (b)

Dolní propust propustí pouze první složku signálu  $x_v(t)$ . Vzhledem k sudosti kosinu dostaneme stejný signál na výstupu pro každý sinusový signál, pro jehož frekvenci platí  $f_R - f_{hDP} < f_s < f_R + f_{hDP}$ . Amplitudová frekvenční charakteristika synchronního detektoru proto odpovídá obr. 5.15 b. Synchronní detektor se tedy chová jako pásmová propust se střední frekvencí  $f_R$  a šírkou  $2f_{hDP}$ . Při měření zašuměného sinusového signálu lze tedy teoreticky neomezeně zvyšovat hodnotu  $S/N$  při snižování  $f_{hDP}$ . Tím se ovšem zpomaluje rychlosť reakce obvodu na změny vstupu. Jde tedy o zvláštní případ frekvenční filtrace. Výhodou je linearita převodní charakteristiky synchronního detektoru. Jde o fázově citlivý usměrňovač schopný zpracovávat libovolně malé signály (na rozdíl od neřízených usměrňovačů s diodami).

Násobením výstupního signálu synchronního modulátoru harmonickým signálem s frekvencí  $f_R$  v druhém obvodu zvaném *synchronní detektor* posuneme signál do původního frekvenčního pásmá. Při detekci je potlačen nejen širokopásmový šum, ale i rušivé signály frekvenčně velmi blízké užitečnému signálu.

V analogových synchronních detektorech je často násobička nahrazena spínačovým obvodem, jehož přenos je po polovinu periody signálu  $x_R(t)$  roven 1 a v druhé polovině periody

<sup>9</sup>Na synchronní detekci lze pohlížet také jako na zvláštní případ korelační filtrace pomocí vzájemné korelační funkce (str. 92). Zpoždění jednoho ze signálů představuje fázový posuv signálu  $x_R(t)$ .

0. Amplitudová frekvenční charakteristika je pak tvořena řadou pásmových propustí na frekvencích  $k \cdot f_R$ ,  $k$ -liché. U číslicových synchronních detektorů se násobení i filtrace realizují algoritmy číslicového zpracování. Referenční harmonický signál se vybírá z tabulky v paměti ROM. Mezní frekvenci dolnofrekvenční propusti lze pohodlně přelaďovat změnou vzorkovací frekvence. Aby se zabránilo aliasingu, musí být frekvenční pásmo signálu omezeno na interval  $(0, f_R)$ . Více o synchronní detekci viz [9].

#### 5.4.5 Společný popis metod zpracování signálu za přítomnosti šumu

Výše popsané metody lze popsat společným vztahem [53] :

$$R(\tau) = \frac{1}{T} \int_T [s(t) + n(t)] \cdot f(t + \tau) d\tau . \quad (5.158)$$

Ve vztahu (5.158) se váhová funkce  $f(t)$  liší podle metody. Při měření autokorelační funkce je  $f(t) = s(t) + n(t)$ , při měření vzájemné korelační funkce je  $f(t) = K \cdot s(t)$ . Pro jednoduché průměrování je  $f(t) = 1$  a pro frekvenční filtraci je  $f(t) = h(-t)$ , kde  $h(t)$  je impulsní odezva použitého filtru.

#### 5.4.6 Vlnková transformace

Kromě uvedených metod se v posledních letech pro vybírání signálu ze šumu používá metoda vlnkové transformace (angl. *wavelet transform*) [22] o které je krátce pojednáno také v druhé části tohoto skripta. Její podrobnější popis přesahuje rozsah tohoto skripta. Výhodou užití vlnkové transformace proti frekvenční filtraci je schopnost zvyšovat poměr  $S/N$  i při překryvajících se spektrech signálu a šumu. Využití této metody pro zvýšení poměru  $S/N$  sestává ze tří kroků:

1. Transformování signálu pomocí waveletu do waveletové domény.
2. Prahování transformovaného signálu (koeficienty waveletové transformace menší než zvolený práh jsou nastaveny na nulu).
3. Zpětná transformace prahovaného signálu.

Uvedeným postupem dojde k potlačení šumu bez zpomalení skoků užitečného signálu ([20, 69]).

# 6. Úvod části o zpracování obrazů (V. Hlaváč)

## 6.1 Zpracování obrazových signálů, vztah k počítačovému vidění a počítačové grafice

Zpracování obrazu může být chápáno jako součást zpracování (měřených) signálů. Návaznost na první část tohoto skripta věnovanou zpracování jednorozměrných signálů je přímočará. Předmětem zájmu je obraz, což je vícerozměrný signál, se kterým se my lidé s ohledem na naši schopnost vidět střetaváme stále. Z obrazu můžeme zjišťovat např. velikost, polohu nebo průměrny jas vybraného objektu. Příslušná změřená hodnota bude smysluplná, jen když také budeme znát přesnost měření.

*Obraz* zde budeme chápát v intuitivním smyslu, tj. jako obraz na sítnici lidského oka nebo obraz sejmuty TV kamerou. Obraz může být modelován matematicky pomocí spojité skalárni funkce  $f$  dvou nebo tří proměnných a budeme jí říkat *obrazová funkce*. V jednoduchém případě je statický obraz popsán obrazovou funkcí dvou souřadnic  $f(x, y)$  v rovině (obecněji v ploše, např. na přibližně kulové ploše sítnice). Obrazová funkce tří proměnných se použije, buď když se plošné obrazy mění v čase  $t$ , tj.  $f(x, y, t)$  nebo v případě objemových obrazů  $f(x, y, z)$ , např. u tomografu.

Hodnoty obrazové funkce odpovídají některé měřené fyzikální veličině, např. jasu u obrazu z černobílé TV kamery, teplotě u termovizní kamery nebo schopnosti pohlcovat záření v daném místě objemu u rentgenového tomografu atd. Pro skalárni obrazovou funkci dvou proměnných  $f(x, y)$  budeme používat generický název *intenzitní obraz*.

V počítači pracujeme s digitalizovanými obrazy, kde je obrazová funkce  $f(x, y)$  představovaná maticí. Prvky matice jsou obrazové elementy (též pixely z angl. picture element), jejichž hodnota je úměrná množství světelné energie. Plocha vzorku není nekonečně malá. Z hlediska dalšího zpracování digitalizovaného obrazu je obrazový element dále nedělitelná nejmenší jednotka. Často budeme pro jednoduchost používat i termín bod, i když půjde o pixel konečných rozměrů. Pokud budeme mít na mysli abstrakci nekonečně malého bodu, bude to z kontextu zřejmé.

Dovolme si úvodelem ambicióznější pohled na obrazy, než poskytuje zpracování obrazu vykládané v tomto skriptu, které se chápe jako vícerozměrné rozšíření teorie signálů. Širší pohled poskytuje *počítačové vidění*, což je vědní obor, který se technickými prostředky snaží napodobit některé schopnosti lidského vidění. Kromě vlastních obrazových vjemů hraje při využití obrazové informace podstatnou roli inteligence člověka a jeho předchozí zkušenosti. Právě inteligence umožňuje reprezentovat dlouho nabývané znalosti či zkušenosti o okolním světě a využívat je i pro řešení nových úloh. Počítačové vidění je v širším smyslu považováno za součást kybernetiky či umělé inteligence. Je to i proto, že se musí při snaze napodobit vidění živočichů vypořádat s řešením špatně podmíněných úloh, velkou algoritmickou složitostí a neurčitostí.

Interpretace obrazových dat je jádrem porozumění obrazu v počítačovém vidění. Interpretace znamená zobrazení: *pozorovaná obrazová data → model*. Toto zobrazení využívá sé-

mantiky, tj. konkrétního světa, kde obrazy představují jisté objekty jako např. buněčná jádra v mikroskopickém snímku, řeky na družicovém snímku Země, kontrolované součástky v průmyslové výrobě. Porozumění obrazu je potom založeno na znalosti, cílech, tvorbě plánů k jejich dosažení a využití zpětných vazeb mezi různými úrovněmi zpracování. Je zde zřejmá snaha napodobit proces vnímání u člověka a jemu podobný způsob rozhodování na základě informace obsažené v obraze.

Zamysleme se nejprve nad otázkou, proč je porozumění (interpretace) v úlohách počítačového vidění těžké? Příčinou jsou:

1. *Ztráta informace při perspektivním zobrazení* původně trojrozměrné scény do dvojrozměrné obrazové roviny čidla kamery. Všechny body na polopřímce dané bodem  $(x, y)$  v obrazové rovině ve směru od středu promítání se zobrazí právě do bodu  $(x, y)$ . Zpětná úloha, která se snaží odvodit trojrozměrné vlastnosti objektů z obrazu jediné kamery, má tedy nekonečně mnoho řešení. Ve skutečnosti ji lze řešit jen tehdy, je-li možné využít dalších dodatečných znalostí. Projektivita je část geometrie, která se zobrazením z 3D do 2D světa zabývá.
2. *Komplikovaný a nejednoznačný vztah mezi jasem*, který měří kamera (a většina jiných obrazových senzorů), a *tvarem povrchu* 3D objektů ve scéně. Jas bodu závisí na mnoha vlivech (odrazivosti povrchu pozorovaného předmětu, poloze a vlastnostech zdrojů světla, orientaci povrchu vzhledem k pozorovateli). Úloha určení 3D vlastností pozorovaných objektů na základě radiometrických měření je také nedostatečně určená úloha.
3. *Veliké množství obrazových dat*. Běžný černobílý televizní signál je nutné digitalizovat do obrazu o rozměru 512x768 bodů v 256 jasových úrovních, chceme-li zachovat kvalitu původního analogového signálu. Takových obrazů je v televizním signálu za sekundu 25. Vynásobením dojdeme k úctyhodnému toku zpracovávané informace, a to 75 megabitů za sekundu. Dnešní nejrychlejší signálové procesory s televizním obrazem v reálném čase nezvládají jiné než nejjednodušší operace. Ani v čase obrazy neobsahují málo dat. Jedna stránka formátu A4 digitalizovaná černobíle skenerem v rozlišení 300 dpi (angl. dots per inch, bodů na palec) a ve 256 jasových úrovních zabírá zhruba 8.3 megabajtů paměti.
4. *Šum*, který je v obraze reálné scény vždy přítomen. Proto je při zpracování velmi často nutné použít pravděpodobnostní techniky, a tedy komplikovanější matematické nástroje. Často ovšem není k dispozici tolik obrazů, aby bylo možné korektně odhadnout statistické vlastnosti obrazových signálů. Nezbývá než se uchylovat k různým zjednodušením.
5. *Vztah mezi pozorovaným detailom a zjištovaným celkem*. Algoritmy zpracování obrazu mohou obvykle analyzovat vlastnosti jen části obrazu prostřednictvím malého okénka. Tímto způsobem se těžko zjišťují globální vlastnosti obrazu, o které většinou jde.

*Počítačová grafika* [16, 101] se od počítačového vidění i zpracování obrazů svým přístupem odlišuje. Většinou skládá z jednoduchých elementů (úseček, oblouků) realistické 2D nebo 3D obrazy. Cílem grafiky je zobrazit (vizualizovat) člověku informaci z počítače, často s možností interakce. Podstatné je i to, že grafika se nemusí vyrovnávat s šumem ve vstupních datech.

## 6.2 Reprezentace obrazu a úlohy analýzy obrazu

Úlohou vnímání obrazu počítačem je najít relaci mezi vstupním obrazem a vytvořenými modely reálného světa. Přechodem od vstupního obrazu k modelu se obrazová informace zhušťuje a přitom se stále více využívá znalostí interpretace obrazových dat. Mezi vstupním obrazem a modelem (úplným popisem, interpretací, klasifikací) se definuje několik úrovní reprezentace obrazové informace. Počítačové vidění se potom skládá z návrhu těchto přechodových reprezentací a algoritmů pro jejich vytváření a zavedení vztahů mezi nimi.

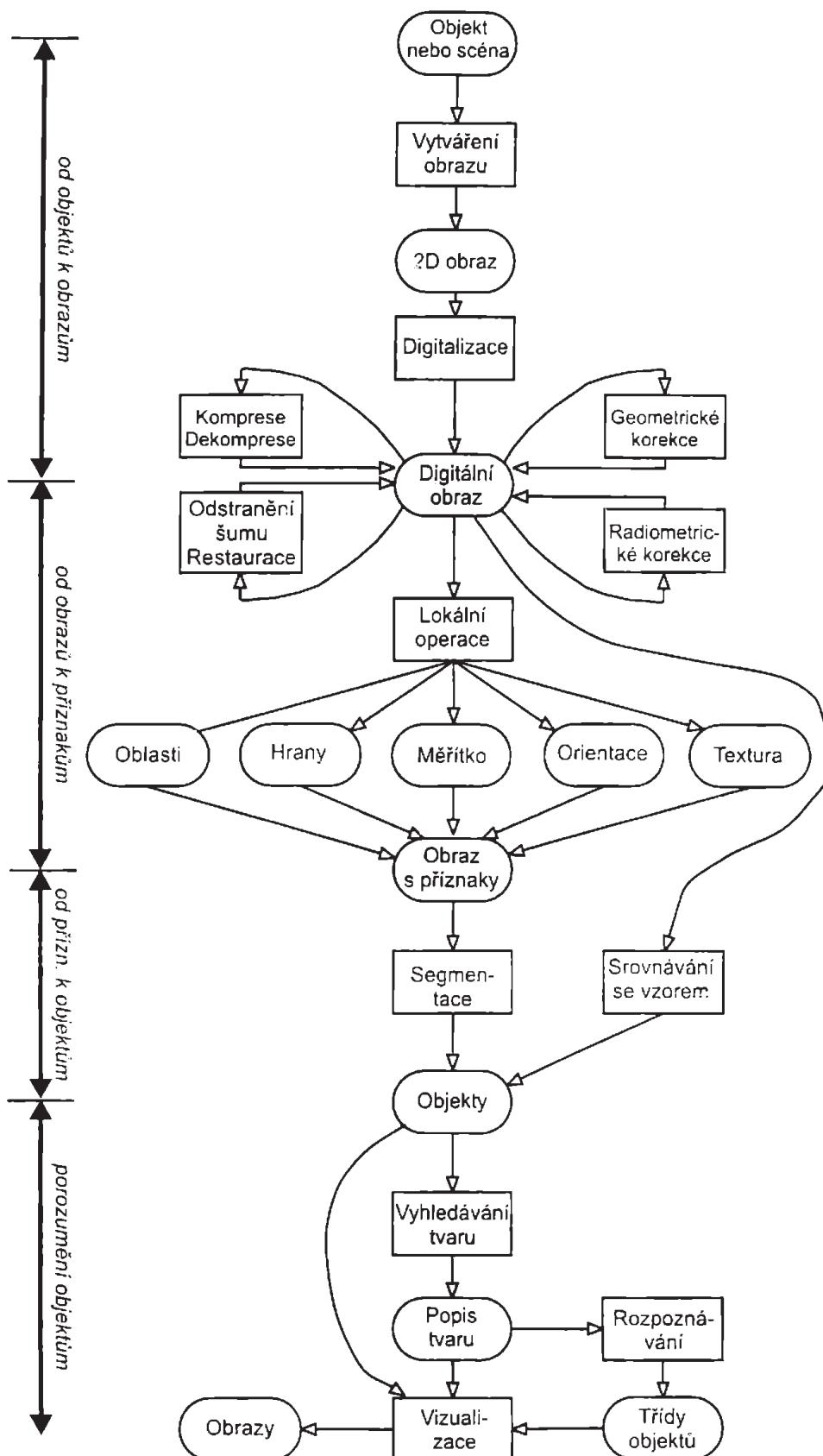
*Reprezentace obrazu* lze hrubě podle organizace dat rozdělit do čtyř úrovní. Hranice mezi nimi však nejsou přesné a v aplikacích se často používá mnohem jemnější odstupňování. I když jsou reprezentace volně seřazeny od signálů nízké úrovně abstrakce k popisu vhodnému pro vnímání, nemusí být tok informace mezi úrovněmi jednosměrný. Zpětné vazby dovolují měnit podle mezivýsledků postup zpracování.

- První, nejnižší *ikonickou* (napodobující) úrovní reprezentace jsou digitální obrazy, které stále mají podobu sejmutých obrazových dat, tj. celočíselné matice s údaji o jasu v příslušných bodech. Takové obrazy jsou i výstupem operací předzpracování (např. filtrace šumu, geometrických nebo radiometrických korekcí), které slouží pro vylepšení obrazu před dalším zpracováním.
- Druhou úrovní reprezentace jsou *příznaky*. Části obrazu jsou spojeny do skupin, které pravděpodobně patří k jednotlivým objektům. Například analýza scény s mnichostěny vyústí buď v hranové segmenty, nebo dvojrozměrné oblasti, které odpovídají stěnám těles. Při hledání příznaků je užitečná informace o konkrétním aplikačním oboru, tedy semantika poskytující možnost obrazové signály alespoň částečně interpretovat. S jejím využitím se daří obejít šum a potíže s částečně chybami obrazovými daty. V příkladu s mnichostěny je touto informací apriorní znalost přínosti hran.
- Třetí úrovní reprezentace jsou *objekty*, které jsou výsledkem segmentace, tj. úplné interpretace obrazových dat. Ve složitějších úlohách se někdy podaří jen částečná segmentace, tj. interpretaci mají jen části obrazů, ale jejich interpretace jako objektů je možná až při dalším, spíše kvalitativním usuzování. V naší ilustraci s mnichostěny by částečnou interpretací mohlo být nalezení stěn mnichostěnu. Úplnou interpretaci se závěrem, že se jedná o šestiboký hranol (např. hlavu šestihranného šroubu), by bylo možné udělat až později.
- Čtvrtou, nejvyšší úrovní reprezentace obrazových dat jsou *relační modely*, které postihují kvantitativní i kvalitativní vlastnosti objektů v obrazu. Zde počítačové vidění využívá technik rozpoznávání a umělé inteligence.

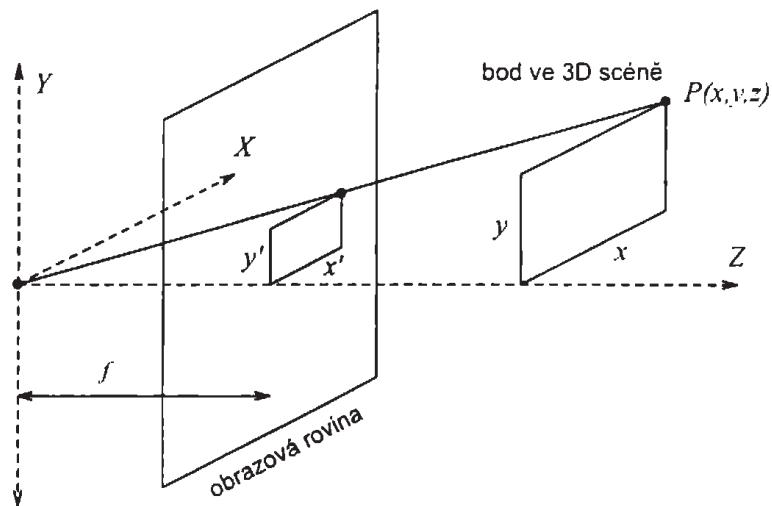
V předmětu Zpracování signálů a obrazů se zabýváme zpracováním obrazu z pozic teorie signálů. Situaci si tudíž proti počítačovému vidění zjednodušíme a zaměříme se jen na počáteční kroky analýzy obrazu, tedy takové, kde se obrazová data neinterpretují nebo se znalost o hledaných objektech využije jen částečně. Stranou také ponecháme případy, kde je cílem porozumět trojrozměrným scénám nebo sekvencím obrazů.

Jednotlivé kroky zpracování dvojrozměrných obrazů vyjádříme jako úlohy, které umožňují splnění cíle analýzy a dovolují přechod mezi jednotlivými reprezentacemi. Abychom získali souhrnnou představu, jak vypadá zpracování a analýza dvojrozměrného obrazu, uvedeme formou blokového diagramu na obr. 6.1 jednotlivé reprezentace (zobrazeny ovály) a příslušné úlohy (zobrazeny obdélníky) pro přechody mezi reprezentacemi. Na obr. 6.1 jsou zachyceny tři ze čtyř dříve zmíněných úrovní reprezentace s hrubším rozlišením a jim příslušné ovály jsou pro názornost šedivě podbarveny.

V tomto skriptu budeme největší díl pozornosti věnovat zpracování obrazu, což na obr. 6.1 odpovídá šesti obdélníkům kolem oválu "digitální obraz". Z lokálních operací budeme podrobněji mluvit jen o detekci hran. Vysvětlíme také, jak digitální obraz vzniká, což odpovídá čtyřem blokům v horní části obr. 6.1.



Obrázek 6.1 Hierarchie úloh zpracování obrazu.



Obrázek 6.2 Geometrie perspektivního promítání.

### 6.3 Obrazová funkce

Svět, který nás obklopuje, je trojrozměrný (3D). Dvojrozměrný (2D) obraz v TV kameře nebo na sítnici oka je výsledkem *perspektivního zobrazení*, neboli středového promítání. Toto zobrazení je pro zjednodušení modelováno *dírkovou komorou*, viz obr. 6.2.

Obrazová rovina v našem modelu na obr. 6.2 je proti dírkové komoře zrcadlově vzhledem k rovině \$xy\$, abychn nedostali obraz se zápornými souřadnicemi. Nechť \$x, y, z\$ jsou souřadnice bodu \$P\$ ve 3D scéně a nechť \$f\$ je vzdálenost obrazové roviny od středu promítání. Poznamenejme, že u objektivů je obrazová rovina umístěna ve vzdálenosti odpovídající ohniskové vzdálenosti. Bod promítнутý do 2D obrazové roviny nechť má souřadnice \$x', y'\$. Potom díky podobným trojúhelníkům pro perspektivní zobrazení platí

$$x' = \frac{xf}{z}, \quad y' = \frac{yf}{z}. \quad (6.1)$$

Perspektivní zobrazení dané předchozím vztahem je nelineární. Někdy se nahrazuje ortografickým neboli paralelním zobrazením, které vznikne, když \$f \rightarrow \infty\$. Prakticky postačuje, když jsou pozorované objekty v mnohem větší vzdálenosti, než je vzdálenost \$f\$ obrazové roviny od středu promítání. Jelikož se všechny body ležící ve 3D prostoru na jedné polopřímce od středu promítání směrem do scény zobrazují do jediného bodu, ztrácí se informace a perspektivní zobrazení není invertovatelné. Proto je geometrická úloha, která má měřit, rekonstruovat či rozpoznat 3D objekty z jejich jediného 2D obrazu, nedostatečně podmíněná. Úlohou 3D rekonstrukce se na FEL ČVUT zabývá až návazný předmět Počítačové vidění a virtuální realita.

Zmínili jsme se již, že vztah mezi jasem měřeným kamerou (který je úměrný hodnotě obrazové funkce) a fyzikálními parametry 3D scény (tvarem objektů daným v každém bodě normálou k povrchu, odrazivostí, zdroji světla, vzájemnou polohou mezi zdroji světla, objektem ve scéně a pozorovatelem) je velmi složitý. Zabývá se jím radiometrie, jejíž základní principy vysvětlíme v sekci 8.1 kapitoly o pořízení obrazu.

V některých vědeckých a technických oborech lze sejmout 2D obrazy, pro které má hodnota obrazové funkce zřejmý fyzikální význam. Uvedme tenký průhledný prosvětlený preparát pozorovaný optickým mikroskopem, písmeno napsané na listu papíru, otisk prstu, obrázek sejmutý plochým skenerem atd.

Základní, často používané a v tomto skriptu popisované postupy analýzy obrazu pracují s 2D obrazem a neodpovídají na složité otázky, které by vysvětlily, jak změřený jas souvisí s fyzikálními vlastnostmi pozorované scény. Aby se úloha zjednodušila, přizpůsobí se snímání např. vhodně voleným osvětlením, a tak se odliší části obrazu zajímavé pro řešenou aplikaci (objekty) od nezajímavých částí (pozadí).

# 7. Digitální obraz

## 7.1 Digitalizace

Čidla pro vstup obrazové funkce jsou většinou zdrojem spojitého signálu. Abychom obrazovou funkci mohli zpracovat v počítači, musíme ji digitalizovat. Stejnou úvalu jsme udělali v kapitole 2. Rozdíl je jen v tom, že obraz je dvojrozměrný. Digitalizace spočívá ve vzorkování obrazu v matici  $M \times N$  bodů a v kvantování spojité jasové úrovně každého vzorku do  $K$  intervalů. Díky kvantování nabývá jasová funkce v digitalizovaných obrazech celočíselných hodnot. Čím jemnější je vzorkování (čím větší  $M, N$ ) a kvantování, tím lépe je approximován původní spojitý obrazový signál.

V souvislosti se vzorkováním spojité obrazové funkce  $f(x, y)$  je potřeba vyřešit dvě otázky. Jednak je třeba určit *interval vzorkování*, což je vzdálenost mezi nejbližšími vzorkovacími body v obrazu. Otázku vzdálenosti vzorků (nebo jinak řečeno plošné vzorkovací frekvence) řeší Shannonova věta o vzorkování, která byla pro jednorozměrný případ vysvětlena v 2.1 a jejíž dvojrozměrné rozšíření je přirozené. Vzorkovací frekvence musí být alespoň dvakrát větší než nejvyšší zajímavá frekvence ve vzorkovaném signálu. Ve zpracování obrazu má Shannonova věta přirozenou fyzikální interpretaci. Interval vzorkování se musí volit tak, aby byl menší nebo rovný polovině rozměru nejmenších detailů v obrazu. Při zpracování obrazů je rozumné vzorkovat alespoň 5-krát jemněji než je teoretická mez dána vzorkovací větou. V případech, kdy je potřeba šetřit pamětí, jako např. při digitalizaci tenkých čar na technických výkresech, nezbývá než vzorkovat různě, a to na teoretické mezi, kde čáře odpovídají dva obrazové body, což je ovšem na úkor robustnosti.

Druhou otázkou je *výběr vzorkovací mřížky*, tj. plošného uspořádání bodů při vzorkování. Obvykle se používá pravidelná mřížka. Existují jen tři pravidelné mnahoúhelníky, jejichž síť úplně pokrývá rovinu, a to rovnostranné trojúhelníky, čtverce a pravidelné šestiúhelníky. V praxi se nejvíce používá čtvercová mřížka, i když je příčinou problémů se spojitostí oblastí, jak uvidíme později při zavádění vzdálenosti.

Jednomu vzorkovacímu bodu odpovídá v digitalizovaném obrazu obrazový element, *pixel*. Po uspořádání do vzorkovací mřížky pokrývají pixely celý digitalizovaný obraz.

Další součástí procesu digitalizace je *kvantování obrazové funkce*. Amplituda ve vzorkovaném obrazu musí být pro zpracování počítačem vyjádřena jako digitální údaj. Počet kvantovacích úrovní má být dostatečně velký, aby byly přesně vyjádřeny jemné detaily obrazu, nevznikaly falešné obrysy a aby se citlivost zařízení blížila citlivosti lidského oka.

Většina systémů pro digitální zpracování obrazu používá kvantování do  $k$  stejných intervalů. Jestliže je pro reprezentaci informace o obrazovém elementu použito  $b$  bitů, je počet úrovní jasu  $k = 2^b$ . Obvykle se používá 8 bitů na obrazový element, někdy 6, jindy postačí 4 bity. Binární obrazy reprezentují informaci o obrazovém bodě jediným bitem. Jen ve specializovaných měřicích systémech se jeden vzorek obrazu reprezentuje 12 a více bitů. Největším problémem při kvantizaci je vznik falešných obrysů u obrazů kvantovaných do nedostatečného počtu jasových úrovní. Tento jev se stane člověku patrný pro počet jasových úrovní menší než

50, což je přibližný počet úrovní jasu, který je člověk schopen v monochromatickém obrazu odlišit. Situaci lze poněkud zlepšit nelineárním kvantováním, které zvětšuje rozsah těch intervalů jasu, jejichž zastoupení není v obrazu pravděpodobné, ale v praxi se používá zřídka.

## 7.2 Vlastnosti digitálního obrazu

V tomto skriptu se budeme především zabývat statickým monochromatickým obrazem popsaným obrazovou funkcí  $f(x, y)$ . Jejími argumenty jsou dvě souřadnice v rovině. Pokud budeme mluvit o multispektrálních, barevných nebo v čase se měnících obrazech, bude to z kontextu zřejmé.

Při zpracování obrazů počítačem se pracuje s digitálními obrazovými funkcemi, které jsou reprezované maticemi. Jejich souřadnice i hodnoty jsou celočíselné. Definičním oborem obrazu je rovinná oblast  $R$

$$R = \{(x, y), 1 \leq x \leq x_m, 1 \leq y \leq y_n\}. \quad (7.1)$$

kde  $x_m, y_n$  jsou maximálními souřadnicemi v obrazu. Obrazová funkce má omezený obor hodnot. Toho lze využít při sumaci nebo integraci (při Fourierově a podobných transformacích), kdy lze použít nekonečné meze při předpokladu nulovosti hodnot obrazu mimo definiční obor  $R$ . Orientaci souřadnic lze zvolit např. v obvyklém kartézském smyslu (osa  $x$  je vodorovná rostoucí vpravo, osa  $y$  svislá rostoucí vzhůru) nebo jako v maticovém počtu, tj. (*řádek, sloupec*).

Obor hodnot obrazové funkce (jasu) je také omezený. Podle konvence odpovídá v monochromatickém obrazu nejnižší hodnota černé a nejvyšší bílé. Mezilehlé hodnoty odpovídají různým stupním šedi.

Kvalita digitálního obrazu je úměrná plošnému, spektrálnímu, radiometrickému a časovému rozlišení. *Plošné rozlišení* je dánno vzdáleností vzorkovacích bodů, což po vytvoření Fourierovského spektra odpovídá také *spektrálnímu rozlišení*. *Radiometrické rozlišení* odpovídá počtu kvantizačních úrovní, tj. intuitivně řečeno počtu hodnot jasu. U časově proměnných obrazů mluvíme ještě o *časovém rozlišení* daným časovým intervalom mezi sejmoutími dvou následujících obrazů.

Matematické transformace pro zde probírané signály předpokládají "rozumnou" obrazovou funkci  $f(x, y)$ , tedy integrovatelnou, mající přímou i inverzní Fourierovu transformaci, atd. Zájemce o existenční problémy Fourierovy transformace speciálních signálů (konstanta, impulsy, neperiodické funkce) odkazujeme na [70]; pro diskrétní obrazy Fourierova transformace existuje vždy.

## 7.3 Metrické a topologické vlastnosti obrazu

Nyní se zabývejme digitálním obrazem z hlediska jeho metrických a topologických vlastností. Uvažujeme digitální obraz složený z pixelů konečných rozměrů uspořádaných do čtvercové vzorkovací mřížky. Některé pojmy, jejichž význam je ve spojité matematice intuitivně zřejmý, nemají pro diskrétní obrazy snadnou analogii. Příkladem je vzdálenost mezi dvěma obrazovými body. Nejdříve si připomeňme, jaké vlastnosti musí mít funkce, abychom jí mohli říkat *vzdálenost* (metrika). Nechť  $p, q, r$  jsou tři body (např. v rovině, 3D prostoru). Funkce  $D$  je vzdáleností, právě když

$$\begin{aligned} D(p, q) &\geq 0, & \text{speciálně } D(p, p) &= 0 \text{ (identita).} \\ D(p, q) &= D(q, p), & & \text{(symetrie).} \\ D(p, r) &\leq D(p, q) + D(q, r), & & \text{(trojúhelníková nerovnost).} \end{aligned}$$

V digitálním zpracování obrazů se používá několik definic vzdálenosti. Mějme dva obrazové body se souřadnicemi  $(x, y)$ ,  $(h, k)$ . *Euklidovská vzdálenost*  $D_E$  je známa z elementární geometrie a každodenní zkušenosti,

$$D_E((x, y), (h, k)) = \sqrt{(x - h)^2 + (y - k)^2}. \quad (7.2)$$

Její výhodou je názornost. Nevýhodou je složitější výpočet kvůli odmocnině a skutečnost, že výsledkem je neceločíselná hodnota, což neodpovídá kvantovaným hodnotám v digitálních obrazech.

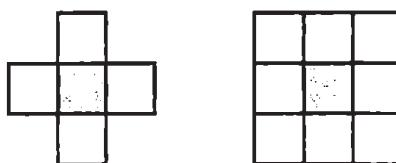
Vzdálenost mezi dvěma body si lze také představit jako nejmenší počet kroků jednotkové vzdálenosti nutných pro přesun z výchozího do cílového bodu v pravoúhlé diskretizační mřížce. Jsou-li povoleny jen kroky o jedno pole v mřížce buď ve vodorovném nebo ve svislém směru, získáme vzdálenost  $D_4$ , které se říká *vzdálenost v městských blocích* díky analogii se vzdáleností ve městě s pravoúhlou sítí ulic,

$$D_4((x, y), (h, k)) = |x - h| + |y - k|. \quad (7.3)$$

Když navíc připustíme i jednotkové kroky v diagonálním směru, je odpovídající vzdáleností  $D_8$ , které se také říká *šachovnicová vzdálenost*, protože odpovídá nutnému počtu kroků krále na šachovnici,

$$D_8((x, y), (h, k)) = \max\{|x - h|, |y - k|\}. \quad (7.4)$$

Pro digitální obrazy je dalším potřebným pojmem *sousednost* pixelů. Dva pixely jsou *4-sousedy*, pokud je jejich vzdálenost  $D_4 = 1$ . Podobně jsou dva pixely *8-sousedy*, když  $D_8 = 1$ . Znázorněme 4-sousedství a 8-sousedství na obr. 7.1.



Obrázek 7.1 Okolí okamžitého obrazového elementu.

Často budeme potřebovat množinu pixelů navzájem vázáných relací sousedství, které se říká *oblast*. Zná-li čtenář teorii množin, můžeme jen říci, že oblast je souvislá (kompaktní) množina. Pokud bychom chtěli pojmem oblasti přece jen vysvětlit, definovali bychom cestu z pixelu  $P$  do pixelu  $Q$  jako posloupnost pixelů  $A_1, A_2, \dots, A_n$ , kde  $A_1 = P$ ,  $A_n = Q$ , přičemž  $A_{i+1}$  je sousedem  $A_i$ ,  $i = 1, \dots, n - 1$ . Potom je oblast množinou pixelů, mezi jejímiž každými dvěma body existuje cesta, která patří celá do oblasti.

Dva pixely v obraze, mezi nimiž existuje cesta, se nazývají *souvislé pixely*. Alternativně bychom mohli říci, že oblast tvoří takové pixely, jejichž každá dvojice je souvislá. Relace "být souvislým" je reflexivní, symetrická a transitivní. Definuje tudíž rozklad množiny (v našem případě obrázku) na třídy ekvivalence (oblasti).

Předpokládejme, že  $R_i$  jsou nesouvislé oblasti, které byly vytvořeny rozkladem obrazu na třídy ekvivalence indukovanými díky relaci *být souvislý*. Abychom se vyhnuli singulárním případům, předpokládejme, že se oblasti nedotýkají okrajů obrazu, jimiž jsou řádky a sloupce v obrazové matici s nejmenšími a nejvyššími indexy. Nechť vznikne oblast  $R$  jako sjednocení oblastí  $R_i$ . Za této předpokladu je užitečné zavést množinu  $R^C$ , která je množinovým doplňkem oblasti  $R$  vzhledem k obrazu. Podmnožina množiny  $R^C$ , která je souvislá s okrajem obrazu se

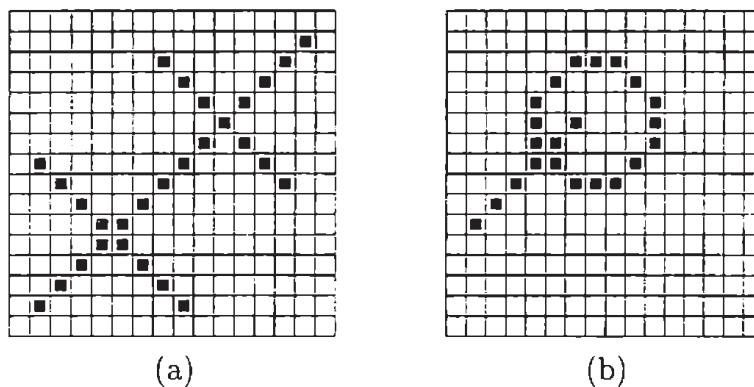
obvykle nazývá *pozadí*, a zbytek komplementu  $R^C$  se nazývá *děrami*. Pokud oblast neobsahuje díry, potom se nazývá *jednoduše souvislou*. Oblasti s děrami se říká *vícenásobně souvislá*.

Všimněte si, že pro zavedení pojmu oblast jsme potřebovali pouze relaci souvislosti. K oblasti lze připojit vlastnosti (atributy), jejichž hodnoty mají původ v interpretaci obrazu. Obvyklé je nazývat některé oblasti obrazu *objekty*. Postup udávající, která oblast bude chápána jako objekt v příslušném světě (pravděpodobně zjednodušeném proti realitě), se nazývá *segmentace obrazu*. V jednoduchém prakticky zajímavém případě, když má bod jas větší než určitý práh, prohlásíme ho za bod jednoho z objektů. Ty body, které patří jedné oblasti, tvoří objekt. Díry jsou potom body, které nepatří objektům a jsou jimi obklopeny. Všechny zbylé body tvoří pozadí. Objekty můžeme označit např. jedničkami a pozadí i díry nulami.

Poznamenejme, že v literatuře se poměrně často pojmem pozadí rozumí sjednocení pozadí a děr v námi definovaném významu. Vyplývá to z úlohy, ve které je potřeba odlišit všechny objekty, které se liší (např. jasem) od zbytku obrazu.

Jas pixelu je jednoduchou vlastností, podle které lze určit objekty v některých obrazech, kde se objekty liší od pozadí právě jasem. Například za body objektu jsou prohlášeny všechny pixely, jejichž jas je větší (nebo v jiných případech menší) než předdefinovaný práh. Příkladem je například tmavý text na bílém papíře, kde jsou za objekty považována jednotlivá písmena. Bílé oblasti obklopené písmenem, např. oblast uvnitř písmene "o". Ostatní bílé části na papíře jsou pozadím.

Sousednost a souvislost definovaná na diskrétní čtvercové mřížce výše uvedeným způsobem vede k paradoxům. Představme si úsečku v digitálním obrazu se sklonem  $45^\circ$ . Při uvažování 4-okolí je však v každém obrazovém elementu nesouvislá. Protíná-li úsečku jiná úsečka, může nastat situace, viz obr. 7.2a, že budí mají (průsečík vlevo dole), nebo nemají (průsečík vpravo nahoře) společný bod.

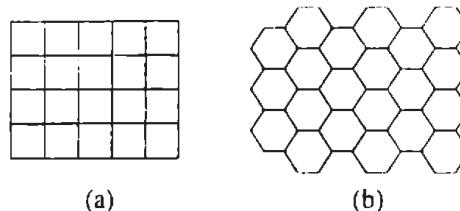


**Obrázek 7.2** (a) Diskrétní úsečky. (b) Paradox uzavřené křivky.

Z euklidovské geometrie je známo, že uzavřená křivka (např. kružnice) dělí rovinu na dvě nesouvislé části. V obrazech digitalizovaných ve čtvercové mřížce lze ve čtvercovém rastru z "vnitřku" uzavřené křivky na obr. 7.2b vést souvislou čáru do "vnějšku", aniž tato protne uzavřenou křivku (nemá s ní společný bod). To však znamená, že "vnitřek" i "vnějšek" uzavřené křivky jsou jedinou oblastí, neboť mezi nimi existuje cesta. To je dalším paradoxem.

Možné řešení nesnáší vyplývajících z uvedeného paradoxu spočívá v tom, že pro objekty použijeme 8-okolí a pro pozadí 4-okolí. Je to však řešení nepřirozené a někdy dosti nepraktické. Exaktnější rozbor paradoxu a jeho řešení pro obrazy s více než dvěma jasovými úrovněmi lze najít v [75, 34].

Uvedené nesnáze, které jsme ukázali na paradoxu souvislosti, jsou příznačné pro čtvercovou diskrétní mřížku. Hexagonální mřížka (obr. 7.3) mnohé z nich řeší. Všech šest sousedů



Obrázek 7.3 (a) Čtvercová mřížka. (b) Hexagonální mřížka.

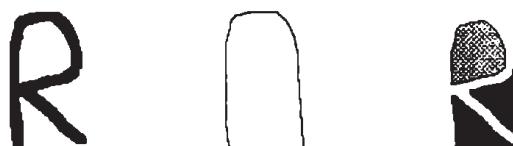
obrazového elementu má stejné vlastnosti, a to především vzájemnou vzdálenost. Širšímu rozšíření hexagonálního rastru brání dva důvody. Za prvé většina digitalizačních zařízení preferuje čtvercový rastr. Za druhé se hexagonální rastr příliš nehodí pro některé operace jako např. Fourierovu frekvenční filtraci.

Alternativním a obecnějším přístupem k řešení problémů souvislosti ve čtvercové mřížce je použití výsledků diskrétní topologie opírajících se o celulární komplexy [50, 51]. Je použit systematický způsob kódování a segmentací obrazů, oblastí a jejich hranic. Rozvíjí se Riemannova myšlenka z 19. století, tj. využít pro reprezentaci struktury různé dimenze: body o dimenzi 0, úsečky hranic o dimenzi 1, plochu pixelu o dimenzi 2, atd. Výhodou je, že je přístup použitelný i pro obrazy s více jasovými úrovněmi.

## 7.4 Další pojmy

Důležitým pojmem v analýze obrazu je *hranice oblasti*. Hranice oblasti  $R$  je množina všech obrazových elementů oblasti, z nichž každý má alespoň jednoho souseda, který nepatří do oblasti  $R$ . Hranice oblasti odpovídá intuitivní představě bodů na samém okraji oblasti. V matematické abstrakci, kterou poskytuje spojitá obrazová funkce, je hranice nekonečně tenká. V digitálním obrazu, pokud je hranice chápána jako složená z pixelů, je třeba rozlišit vnitřní a vnější hranici. *Vnitřní hranice* potom odpovídá předchozí definici hranice. *Vnější hranice* by byla hranicí pozadí, tj. množinového doplňku k oblastem. Poznámenejme, že k lepší definici hranice vede zmíněná topologie zavedená pomocí celulárních komplexů [49]. Odpovídá lépe představě zavedené ve spojité matematice, kde má hranice nulovou tloušťku.

Pro popis tvaru objektů (oblastí) se také používá konvexní obal. *Konvexní obal* objektu je nejmenší oblast obsahující objekt taková, že každé dva body oblasti mohou být spojeny úsečkou, jejíž všechny body patří do oblasti. Ukažme si konvexní obal na příkladě objektu, jehož tvar připomíná písmeno R, viz obr. 7.4, vlevo. Pro názornost si představme, že na objekt natáhneme gumičku. Konvexní obal potom bude oblast, jejíž hranice bude dána gumičkou, viz obr. 7.4, uprostřed.



Obrázek 7.4 Popis pomocí topologických složek: objekt tvaru písmene R, jeho konvexní obal, jezera a zálivy.

Množině bodů uvnitř konvexního obalu, které objektu nepatří, se říká deficit konvexnosti objektu. Existují dva druhy deficitu konvexnosti. První tzv. *jezera* (na obr. 7.4 vpravo šrafováno) jsou plně ohrazena objektem. Druhé tzv. *zálivy* (na obr. 7.4 vpravo vyplněno) leží mezi

konvexním obalem a objektem. V některých aplikacích se objekty popisují pomocí konvexního obalu, jezer a zálivů.

Pro diskrétní mřížku musíme poněkud modifikovat význam konvexního obalu, aby si definice zachovala předchozí smysl. Objekty jako trojúhelníky a kružnice je třeba považovat za konvexní, i když jejich hranice mají schodovitý charakter. Algoritmy pro výpočet konvexního obalu lze najít v [89].

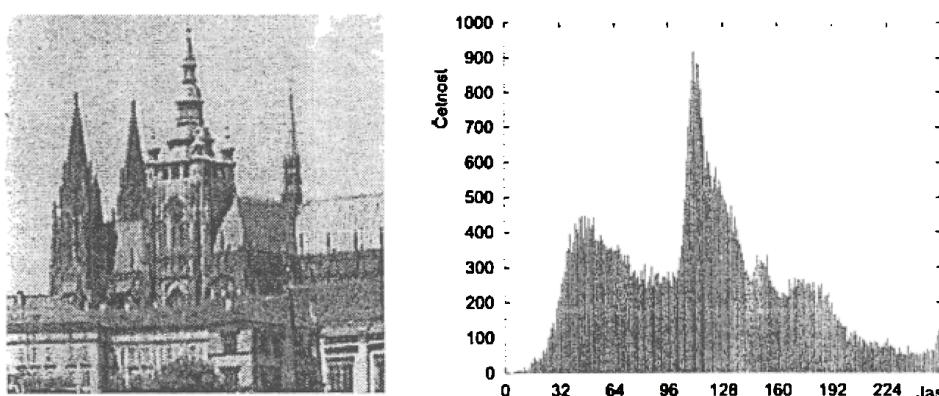
Matematická disciplína zvaná topologie dovoluje studovat vlastnosti, které se neopírají o pojem vzdálenosti. Jinak řečeno jsou invariantní k *homeomorfní transformaci*, kterou pro případ obrazů můžeme připodobnit k transformaci "gumové plochy". Lze si ji představit jako roztahování gumového pouťového balónku, na němž je nakreslen objekt. Roztahováním balónku ale nezměníme souvislost částí nakresleného domečku, nezmění se ani počet dér v oblastech. Jednou ze základních topologicky invariantních vlastností obrazu je Euler-Poincarého charakteristika, která se určí jako počet oblastí minus počet dér.

Vlastnosti probírané dosud v této sekci se opírají o pojem oblasti, což předpokládá, že víme, které pixely obrazu k oblasti přiřadit. Jinak řečeno, očekává se, že umíme obraz interpretovat (neboli segmentovat). Na počátku zpracování ale obvykle obraz segmentovat neumíme a obrazová funkce  $f(x, y)$  představuje dvourozměrný signál. Hledají se takové nástroje, které ho umějí popsat bez nutnosti interpretace. K tomu se často používají *lokální operace* využívající jen bezprostřední okolí právě zpracovávaného pixelu. Pro přiblížení si představme, že se lokální operace musí spokojit s podobnou informací, jako kdybyste se na fotografiu mohli dívat jen velmi malým otvorem v neprůhledném papíru položeném na fotografii.

Významnou lokální informací o obrazu lze získat pomocí hran, které udávají, jak se lokálně mění obrazová funkce. V matematice o funkci analogicky vypovídá její první derivace (gradient). *Hrana* je tedy vlastností obrazového elementu a jeho bezprostředního okolí. Hrana odpovídá *gradientu* obrazové funkce  $f(x, y)$ . Jelikož je funkce  $f(x, y)$  dvojrozměrná, je hrana určena velikostí a směrem. *Velikost* hrany odpovídá modulu gradientu spojité obrazové funkce v příslušném pixelu. *Směr* hrany je kolmý na směr gradientu, který míří ve směru největšího růstu obrazové funkce. O metodách hledání hran je podrobněji pojednáno v 11.4.

Proč definujeme směr hrany natočený vzhledem ke směru gradientu? Na začátku této sekce byla definována hranice oblasti. Příčinou natočení je možnost řetězit při segmentaci hrany, spojovat je a získávat tak hranice oblasti.

Představu o rozdělení jasových úrovní v digitálním obrazu získáme pomocí histogramu. *Histogram jasu*  $h_f(z_i)$ ,  $i = 0, \dots, L - 1$  je vektor  $h$  s počtem složek rovným počtu jasových úrovní  $L$ . Hodnota každé složky odpovídá četnosti bodů příslušného jasu v obrazu. Histogram lze chápát jako odhad hustoty pravděpodobnosti rozdělení jasu v obrazu.



Obrázek 7.5 Histogram jasu pro  $L = 256$  jasových úrovní.

Histogram se často zobrazuje jako sloupcový diagram. Příklad obrazu a jeho histogramu je uveden na obr. 7.5.

Obraz lze také chápat jako realizaci náhodného procesu. Mohlo by nás také zajímat, jaké je rozdělení pravděpodobnosti (též hustota prvního řádu)  $p_1(z, x, y)$ , která udává pravděpodobnost, že pixel  $(x, y)$  má jas  $z$ . Pokud nás nezajímá pozice v obrazu, získáme hustotu  $p_1(z)$ , jejímž odhadem je histogram jasu.

Histogram velmi často bývá jedinou globální informací o obrazu. Můžeme ho použít při nastavování podmínek pro snímání a digitalizaci, při změnách jasové stupnice a při segmentaci obrazu na objekty a okolí. Určitému obrazu je přiřazen jeden histogram. Jednomu histogramu však může odpovídat několik obrazů. Například při změně polohy objektu na pozadí o konstantním jasu se histogram nezmění.

Histogram digitalizovaného obrazu má řadu lokálních minim a maxim, která mohou komplikovat jeho další použití. Proto se histogram jasu někdy před dalším použitím vyhlazuje, aby se potlačily jeho lokální poruchy. Vyhlazování je založeno na vyhlazení sousedních prvků pole s histogramem. Nejjednodušší filtrací je klouzavý průměr. Nový histogram  $h'_f(z_i)$  vypočte jako

$$h'_f(z_i) = \frac{1}{2K+1} \sum_{j=-K}^{K} h_f(z_{i+j}), \quad (7.5)$$

kde konstanta  $K$  udává velikost okolí.

## 7.5 Barevný obraz

Obraz může být v jednodušším případě monochromatický. Je reprezentován jedinou obrazovou funkcí  $f(x, y)$ . Ve složitějším případě pracujeme s barevným (multispektrálním) obrazem. Každé dvojici plošných souřadnic  $(x, y)$  odpovídá vektor hodnot, např. jasů pro jednotlivé barevné složky obrazu. Barva je vlastnost objektů spojená s jejich schopností odrážet elektromagnetické vlnění různých vlnových délek. Při snímání a počítačovém zpracování nelze pracovat se všemi vlnovými délками obrazu. Oblast vlnění zajímavou z hlediska zpracování se snažíme rozdělit do několika spektrálních pásem, která pokrývají požadovaný rozsah vlnových délek. Obraz potom snímáme více senzory, z nichž každý je citlivý na poměrně úzkou část spektrálního pásma.

Výběr vlnových délek pro jednotlivá spektrální pásma závisí na aplikaci. Multispektrální snímky se často používají v dálkovém průzkumu Země z leteckých a kosmických nosičů. Např. družice LANDSAT 4 (USA) vysílá digitalizované snímky v pěti spektrálních pásmech od blízkého ultrafialového až po infračervené záření.

Barevný signál v normě barevné televize se skládá ze tří samostatných spektrálních složek RGB (červená má vlnovou délku 700 nm, zelená 546,1 nm, modrá 435,8 nm). Smícháním složek vzniká barevný obraz. Každá složka se digitalizuje, přenáší a případně zpracovává počítačem samostatně. Existují i algoritmy využívající vzájemné závislosti mezi jednotlivými barevnými složkami.

Digitální zpracování multispektrálních a barevných obrazů většinou používá obdobné metody jako při práci s monochromatickými obrazy. Postupy se aplikují na jednotlivé spektrální složky. Při vyhodnocení multispektrálních a barevných obrazů je třeba vzít v úvahu dva aspekty: velké množství dat a značnou podobnost mezi odpovídajícími daty ve složkách.

# 8. Pořízení obrazu

## 8.1 Obrazy jako radiometrická měření

Pro zobrazování objektů lze využít tři druhy vyzářené energie

1. Elektromagnetické záření včetně  $\gamma$  záření, rentgenového záření (angl. X rays), ultrafialového záření, viditelného světla, infračerveného záření, mikrovln a radiových vln. Ve vakuu se záření šíří rychlostí světla. Látkami se záření šíří pomaleji a navíc v závislosti na vlnové délce. V tomto skriptu se soustředíme pouze na viditelnou část elektromagnetického záření.
2. Záření častic, např. elektronů a neutronů.
3. Akustických vln v plynech, kapalinách a tuhých tělesech. V plynech a kapalinách se šíří pouze podélná vlna a v tuhých látkách se přidává i příčná vlna. Rychlosť šíření akustických vln je přímo úměrná elastickým vlastnostem média, kterým prospupuje.

Záření interaguje s hmotou, a to buď na povrchu objektů nebo v jejich objemu. Záření je z objektů vyzařováno buď díky vlastnímu tepelnému pohybu molekul (horké zářící těleso) nebo díky vnější stimulaci (např. odražené záření, luminiscence, aj.).

Záření doprovází informace, kterou lze využít k identifikaci viděných objektů a měření jejich některých vlastností. Jde o:

1. *Frekvenci* záření nebo jinak vyjádřeno vlnovou délku.
2. *Amplitudu*, tj. intenzitu.
3. *Mód polarizace* pro příčné vlny.
4. *Fázi*, která je ale přístupná pouze při použití koherentních zobrazovacích technik, jakými je interferometrie nebo holografie.

Dále vysvětlíme otázku vzniku obrazu díky odrazu záření od povrchů neprůhledných objektů ve spektru viditelného světla, a to z radiometrického hlediska.

### 8.1.1 Radiometrické úvahy, pojmy

TV kamera a většina ostatních obrazových senzorů měří množství světelné energie dopadající na jednotlivé pixely. Výsledkem je intenzita, které se také často říká jas. *Radiometrie* je část fyziky, která se zabývá tokem a přenosem vyzářené energie. Z našeho hlediska radiometrie dovoluje vysvětlit mechanismus vzniku obrazu. Neformálně řečeno, jas v daném pixelu závisí na tvaru objektu, odrazivých vlastnostech jeho povrchu, poloze pozorovatele a poloze a typu světelných zdrojů.

V praktickém počítačovém vidění a aplikacích se lidé velmi často nesnaží pochopit vznik a fyzikální význam hodnot obrazové funkce (jasu). Je to pro malou praktickou použitelnost výsledků takové analýzy, a to díky její složitosti, silně omezujícím podmínkám na odrazivé vlastnosti materiálu a numerické nestabilitě dostupných metod. Údaj o změřené jasové hodnotě u běžných levnějších kamer je spíše kvalitativní než kvantitativní, protože radiometrická přesnost takových kamer je mnohem menší než geometrická a hlavně je vztah mezi měřeným

jasem a tvarem příliš složitý. Cestou, jak tyto problémy obejít, je použít osvětlení scény na vržené právě pro řešenou úlohu a pokusit se odlišit objekty od pozadí. Podaří-li se to, řeší se řádově jednodušší úloha, ovšem za cenu ztráty informace, kterou by radiometrická analýza změn jasu mohla přinést.

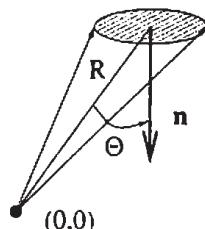
Disciplínou spřízněnou s radiometrií je *fotometrie*, která využívá veličiny popisující odezvu vjemů zrakového smyslu u člověka. Fotometrické veličiny závisí na spektrální charakteristice zdroje záření a citlivosti světločivých buněk na sítnici oka. Radiometrie i fotometrie popisují podobné jevy a používají k tomu podobné veličiny. V této kapitole budeme uvádět jednotky fyzikálních veličin v hranatých závorkách. Kde by mohlo dojít k záměně mezi radiometrickými a fotometrickými veličinami, označíme fotometrické dolním indexem  $_{ph}$  a radiometrické necháme bez indexu.

Základní radiometrickou veličinou je *zářivý tok*  $\Phi$  [W] a jeho fotometrický protějšek *světelný tok*  $\Phi_{ph}$  [lm (=lumen)]. Mezi oběma veličinami platí převodní vztah  $1 \text{ [W]} = 680 \text{ [lm]}$ , ale pouze pro vlnovou délku světla  $\lambda = 555\text{nm}$  a fotooptické (čípkové) vidění [33]. Schopnosti vnímat se u různých lidí liší. Proto byl ve fotometrii mezinárodní organizací CIE (francouzsky Commission Internationale de l'Eclairage) definován standardní fotometrický pozorovatel, který odpovídá statisticky zjištěným průměrným lidským schopnostem.

Nechť  $K(\lambda)$  je *světelná účinnost* [ $\text{lm W}^{-1}$ ],  $S(\lambda)$  [W] je výkon příslušného zdroje záření,  $\lambda$  [m] je vlnová délka. Potom světelný tok  $\Phi_{ph}$  udává celkovou vnímanou odezvu vjemu

$$\Phi_{ph} = \int_{\lambda} K(\lambda)S(\lambda)d\lambda. \quad (8.1)$$

Jelikož fotometrické veličiny závisí na pozorovateli, budeme se snažit výklad opírat spíše o radiometrické veličiny. Úlohy počítacového vidění se tak mohou více opírat o optickou podstatu vnímaných vjemů.



Obrázek 8.1 Prostorový úhel odpovídající elementární ploše.

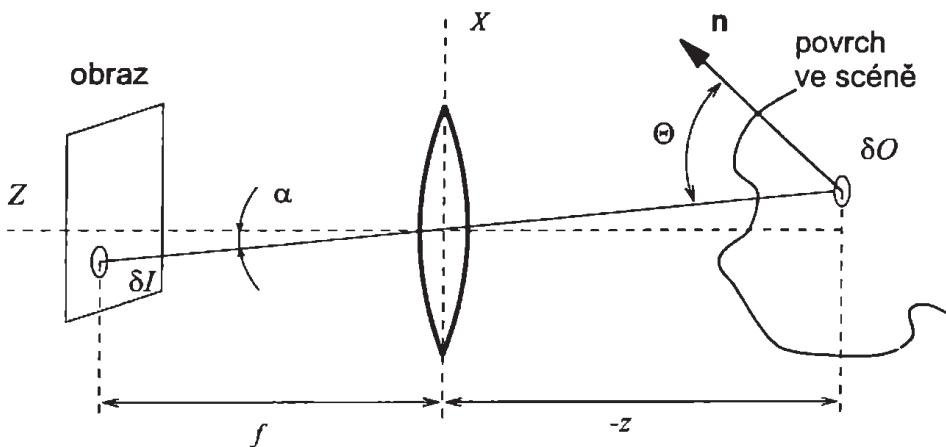
Povrch tělesa může energii odrážet do celé polokoule, a to do různých směrů různě. *Prostorový úhel* je dán plochou na jednotkové kouli, kterou ohraňuje kužel s vrcholem ve středu koule. Celé polokouli bude odpovídat prostorový úhel  $2\pi$  [sr, steradián]. Malá ploška  $A$  ve větší vzdálenosti  $R$  od počátku souřadného systému, tj.  $R^2 \gg A$ , s úhlem  $\Theta$  mezi normálou k ploše a spojnicí polopřímou od počátku k ploše odpovídá prostorovému úhlu  $\Omega$  [sr], viz obr. 8.1.

$$\Omega = \frac{A \cos \Theta}{R^2}. \quad (8.2)$$

*Ozáření* (též intenzita ozáření)  $E$  [ $\text{W m}^{-2}$ ] udává výkon světelné energie, které dopadá na jednotku plochy povrchu tělesa,  $E = \delta\Phi\delta A$ , kde  $\delta A$  je nekonečně malá ploška na povrchu. Odpovídající fotometrická veličina je *osvětlení* [ $\text{lm m}^{-2}$ ]. Ve zpracování obrazu se obvykle neformálně používá fotometrická veličina *jas* (též měrná svítivost, stupeň šedi)  $L_{ph}$  [ $\text{lm m}^{-2} \text{ sr}^{-1}$ ] pro označení veličiny, kterou měří kamera. Radiometrický ekvivalent jasu je *zář* (též měrná zářivost)  $L$  [ $\text{W m}^{-2} \text{ sr}^{-1}$ ] udává světelný výkon, který se z jednotky plochy povrchu tělesa vyzáří do určitého prostorového úhlu [sr].

### 8.1.2 Rovnice ozáření

Ozáření  $E$  je dáno množstvím energie, kterou senzor snímající obraz získá na jednotku účinné plochy čidla světla [34]. Účinná plocha čidla respektuje vzájemné natočení elementární plošky, ze které je světelná energie vyzařována, a elementární plošky přijímacího čidla. Budeme hledat vztah mezi ozářením  $E$  měřeným v obrazu a září  $L$  malé plošky v pozorované scéně, jejíž energii zachytí čočka. Odvození založíme na zjednodušené představě čidla, např. jednoho pixelu digitální kamery. Geometrické uspořádání je schematicky znázorněno na obr. 8.2. Optická osa je ve směru osy  $Z$ . Objektiv je umístěn v počátku a má ohniskovou vzdálenost  $f$ . Elementární ploška  $\delta O$  na povrchu objektu, jejíž zobrazení do plošky v obrazu  $\delta I$  nás zajímá, je ve vzdálenosti  $z$ . Spojnice mezi ploškami  $\delta O$  a  $\delta I$  a osa  $Z$  svírají úhel  $\alpha$ . Spojnice mezi ploškami  $\delta O$  a  $\delta I$  svírá s normálou  $n$  k ploše  $\delta O$  úhel  $\Theta$ .



**Obrázek 8.2** Vztah mezi ozářením  $E$  a září  $L$  pro kameru realizující perspektivní zobrazení.

Uvažujme perspektivní zobrazení, ve kterém paprsky procházejí počátkem. Paprsek procházející středem čočky se neláme, a proto se prostorový úhel odpovídající elementární ploše ve scéně rovná prostorovému úhlu příslušejícímu elementární ploše v obrazu. Nakloněná ploška, jak je viděna ze středu soustavy, je  $\delta I \cos \alpha$  a její vzdálenost od středu optické soustavy je  $f / \cos \alpha$ . Odpovídající prostorový úhel je

$$\frac{\delta I \cos \alpha}{\left(\frac{f}{\cos \alpha}\right)^2}.$$

Analogicky pro prostorový úhel pro nakloněnou plošku  $\delta O$  na povrchu scény platí

$$\frac{\delta O \cos \Theta}{\left(\frac{z}{\cos \alpha}\right)^2}.$$

Z rovnosti prostorových úhlů plyne

$$\frac{\delta O}{\delta I} = \frac{\cos \alpha z^2}{\cos \Theta f^2}. \quad (8.3)$$

Nyní stanovme, jaké množství světla projde přes čočku, je-li její průměr  $d$ . Prostorový úhel  $\Omega_L$ , kterým je čočka viděna z elementární plošky na objektu, je dán vztahem

$$\Omega_L = \frac{\pi}{4} \frac{d^2 \cos \alpha}{\left(\frac{z}{\cos \alpha}\right)^2} = \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha. \quad (8.4)$$

Nechť  $L$  je zář povrchu objektu otočeného směrem k čočce. Potom elementární příspěvek zářivého toku  $\Phi$  [W] dopadajícího na čočku je

$$\delta\Phi = L\delta O \Omega_L \cos\Theta = \pi L\delta O \left(\frac{d}{z}\right)^2 \frac{\cos^3\alpha \cos\Theta}{4}. \quad (8.5)$$

Světelná energie je čočkou soustředěna do obrazu. Zanedbáme ztráty v čočce a uvažujeme, že žádné další světlo na element obrazu nedopadá. Potom pro ozáření  $E$  elementární plošky obrazu platí

$$E = \frac{\delta\Phi}{\delta I} = L \frac{\delta O}{\delta I} \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3\alpha \cos\Theta. \quad (8.6)$$

Po substituci za  $\frac{\delta O}{\delta I}$  podle rovnice (8.3) dostaneme tzv. *rovnici ozáření*, která říká, jak se v obrazu projeví zář scény

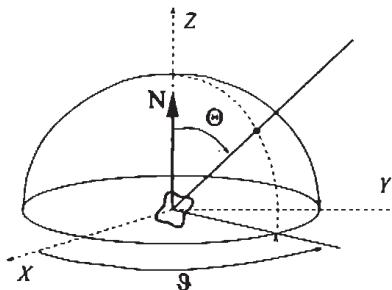
$$E = L \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4\alpha. \quad (8.7)$$

Činitel  $\cos^4\alpha$  popisuje systematickou optickou vadu zvanou *vinětace*, která popisuje jev, kdy jsou paprsky lámající se s větším úhlem  $\alpha$  (dále od optické osy) více zeslabovány. Tato chyba je více patrná u širokoúhlých objektivů než u teleobjektivů. Jelikož vinětace je systematickou chybou, lze ji pro radiometricky kalibrovanou kameru kompenzovat.

Činitel  $n_f = \frac{f}{d}$  je nazýván *f-číslem objektivu* a udává, nakolik se objektiv liší od dírkové kamery. V rovnici (8.7) se objevila převrácená hodnota  $n_f$ .

### 8.1.3 Odrazivost povrchu

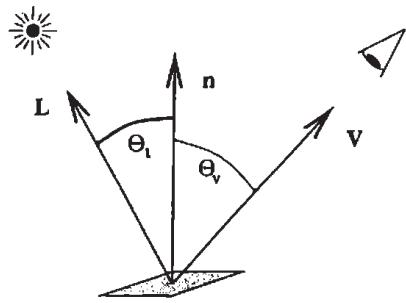
V některých úlohách počítačového vidění a počítačové grafiky je hodnota obrazové funkce chápána jako odhad záře  $L$ , která vznikla odrazem světelné energie od povrchu scény. Proto je potřebné rozumět jevům, které při odrazu vznikají. Uvedeme zde jen stručné principiální vysvětlení. Zájence o podrobnosti odkazujeme na [37, 16, 47].



Obrázek 8.3 Orientace plošky se popisuje v kulových souřadnicích azimutem  $\varphi$  a polárním úhlem  $\Theta$ .

Zář matného objektu, který nevyzařuje vlastní energii, je ovlivněn ozářením povrchu objektu. Místo záře se častěji používá jas. Jas, jak už víme, je září přímo úměrný. Ozáření závisí na typu světelných zdrojů (zda jsou např. bodové či plošné) a jejich umístění vzhledem k ploše a pozorovateli. Jas viděný pozorovatelem je za prvé ovlivněn schopností povrchu odrážet světelnou energii, což se popisuje činitelem odrazivosti. Za druhé je jas určen lokální geometrií plošky, která se udává její normálou  $n$ . Popišme si nejdříve souřadnou soustavu, ve které se geometrie plošky vyjadřuje, viz obr. 8.3. Polární úhel je označen  $\Theta$  a azimut  $\varphi$ .

Nyní se budeme snažit vyjádřit schopnosti různých materiálů odrážet na povrchu objektu světlo. Nechť je pro jednoduchost elementární ploška osvětlena z bodového zdroje, obr. 8.4.



Obrázek 8.4 Popis směrových vektorů ke světelnému zdroji a pozorovateli.

Polární úhel a azimut příslušný bodovému zdroji má index  $i$ . Směr k pozorovateli je označen indexem  $v$ . Označme pomocí  $\delta E(\Theta_i, \varphi_i)$  ozáření plošky, tj. množství světla, které ze světelného zdroje plošku ozařuje. Elementární příspěvek záře z plošky směrem k pozorovateli je  $\delta L(\Theta_v, \varphi_v)$ . V obecném případě se odrazivost těles popisuje dvousměrovou distribuční funkcí odrazu  $f_r$  [ $sr^{-1}$ ] (angl. bidirectional reflectance distribution function), zkracujeme BDRF [66],

$$f_r(\Theta_i, \varphi_i; \Theta_v, \varphi_v) = \frac{\delta L(\Theta_v, \varphi_v)}{\delta E(\Theta_i, \varphi_i)}. \quad (8.8)$$

BDRF  $f_r$  udává jas elementární plošky na povrchu pro určitý materiál, zdroj světla a směr pohledu. Modelování dvousměrové distribuční funkce odrazu  $f$  je důležité pro realistické stínování v počítačové grafice [16]. Například pro povrchy s orientovanou mikrostrukturou, jako má polodrahokam tygří oko (křemen prostoupený vlákny krokydolitu, žlutohnědá barva, měňavý třpyt), duhové opeření některých ptáků (paví pera) nebo např. hrubý řez hliníkem, je nutné modelovat  $f_r$  v plné složitosti dané rovnici (8.8).

Naštěstí pro většinu prakticky významných povrchů není odrazivost popsána dvousměrovou distribuční funkcí odrazu  $f_r$  závislá na otočení podél normály k povrchu. Potom  $f_r$  závisí pouze na rozdílu azimutů směrů ke světelnému zdroji a pozorovateli  $\varphi_i - \varphi_v$ , tedy  $f_r(\Theta_i, \Theta_v, (\varphi_i - \varphi_v))$ . Toto zdjednodušení platí pro matné (lambertovské) povrchy, ideálně zrcadlící povrch a řadu dalších.

$E_i(\lambda)$  označuje zář způsobenou ozářením plošky povrchu scény a  $E_r(\lambda)$  je tok energie na jednotku plochy vyzářený ploškou zpět do poloprostoru. Poměr

$$\rho(\lambda) = \frac{E_r(\lambda)}{E_i(\lambda)} \quad (8.9)$$

se nazývá *koefficient odrazivosti* nebo *albedo*. Albedo vyjadřuje, jaký podíl dopadající energie je odražen zpět do poloprostoru. Předpokládejme pro jednoduchost, že můžeme zanedbat vliv barvy povrchu, a také předpokládejme, že albedo nezávisí na vlnové délce  $\lambda$ . Hledaný podíl je potom integrálem záře z povrchu  $L$  v prostorovém úhlu  $\Omega$ , který odpovídá celému poloprostoru

$$E_r = \int_{\Omega} L d\Omega. \quad (8.10)$$

Nyní zavedeme *funkci odrazivosti*  $R(\Omega)$ , která modeluje vliv lokálních změn geometrie povrchu na rozptýlení odražené energie v prostoru.  $\Omega$  je nekonečně malý prostorový úhel kolem směru pohledu.

$$\int_{\Omega} R d\Omega = 1. \quad (8.11)$$

Obecně odrazivé vlastnosti povrchu závisí na třech úhlech, popisujících vzájemný vztah mezi směrem ke zdroji světla  $\mathbf{L}$ , směrem k pozorovateli  $\mathbf{V}$ , lokální orientaci povrchu danou normálnou  $\mathbf{n}$  (vzpomeňme na obr. 8.4). Kosiny těchto úhlů lze napsat jako skalární součin vektorů. Proto je funkce odrazivosti funkcií tří skalárních součinů

$$R = R(\mathbf{n} \cdot \mathbf{L}, \mathbf{n} \cdot \mathbf{V}, \mathbf{V} \cdot \mathbf{L}) . \quad (8.12)$$

*Lambertovský povrch* (také ideálně matný, ideálně difúzní povrch) odráží světelnou energii rovnoměrně do všech směrů. Proto je zář (a také jas) ze všech směrů konstantní. Dvousměrová distribuční funkce odrazivosti  $f_{\text{Lambert}}$  je také konstantní

$$f_{\text{Lambert}}(\Theta_i, \Theta_v, \varphi_i - \varphi_v) = \frac{\rho(\lambda)}{\pi} . \quad (8.13)$$

Když se předpokládá konstantní albedo  $\rho(\lambda)$ , potom lze odrazivost lambertovského povrchu vyjádřit ve známém tvaru kosinového zákona

$$R = \frac{1}{\pi} \mathbf{n} \cdot \mathbf{L} = \frac{1}{\pi} \cos \Theta_i . \quad (8.14)$$

Všimněte si, že funkce odrazivosti lambertovského povrchu nezávisí směru pohledu  $\mathbf{V}$ . Lambertovský model odrazivosti je pro svou jednoduchost velmi oblíbený.

**Příklad 8.1 Koeficienty odrazivosti pro typické lambertovské povrhy.** Je-li lambertovský povrch osvětlen rovnoběžnými paprsky světla s polárním úhlem  $\Theta$  a ozářením  $E$ , je pozorovaná zář  $L$ . Uvedeme několik příkladů materiálů [24], jejichž odrazivost lze považovat za lambertovskou s hodnotami odrazivosti  $\rho(\lambda)$  pro  $\lambda$  odpovídající asi středu viditelného spektra: bílý piják 0,8, bílý psací papír 0,68, bílý strop nebo žlutý papír 0,6, tmavě hnědý papír 0,14 a tmavý samet 0,004. ▲

*Ideální zrcadlový povrch* odráží ozáření ze směru  $(\Theta_i, \varphi_i)$  do směru  $(\Theta_i, \varphi_i + \pi)$ . Vlastní povrch tedy není vidět, ale ukazuje jen zdánlivý zrcadlově převrácený obraz zdrojů osvětlení.

Nyní udělejme malou odbočku a zavedeme pojem *gradientní prostor*, což je jen jiný formalismus, jak vyjádřit normály k povrchu. Nechť  $g(x, y)$  popisuje výšku povrchu v každém jeho bodě. Potom gradient k povrchu  $s$  je dán

$$\mathbf{s} = \nabla g(x, y) = \left[ \begin{array}{c} \frac{\partial g}{\partial x}, \\ \frac{\partial g}{\partial y} \end{array} \right]^T = [p, q]^T .$$

Gradientní prostor  $\mathbf{s}$  lze také vyjádřit vzhledem k normálám povrchu  $\mathbf{n}$ . Jeho první dvě složky představují záporně vzaté složky prostorového gradientu  $\mathbf{s}$  a složka ve směru osy  $z$  je jednotková

$$\mathbf{n} = [ -p, -q, 1 ]^T .$$

Když jsme zavedli gradientní prostor, můžeme vyjádřit závislost záře objektu  $L$  na lokální orientaci povrchu pomocí *mapy odrazivosti*  $R(p, q)$ , a ta v gradientním prostoru vytváří vnořené isokřivky odpovídající stejné hodnotě odrazivosti. Hodnoty mapy odrazivosti  $R(p, q)$  mohou být:

1. změřeny experimentálně na přístroji zvaném goniometr, kde lze mechanicky nastavit úhly  $\Theta$  a  $\varphi$ . Vzorek měřeného předmětu je umístěn ve středu kulových souřadnic a odrazivost se měří přesným bodovým čidlem pro různé úhly  $\Theta$  a  $\varphi$ ;

2. změřeny kamerou z jednoho snímku objektu homogenně pokrytého měřenou látkou za předpokladu známého tvaru (obvykle polokoule);
3. odvozeny z matematického modelu popisující odrazivé vlastnosti.

Kromě již zmíněného lambertovského modelu odrazivosti pro ideálně matné povrchy se používá Phongův model napodobující odraz dielektrických materiálů, Torrancův-Sparrowův model, který povrch charakterizuje mnoha rovinnými mikropovrchy (fasety) s normálami splňujícími gaussovské rozdělení a konečně Beckmanův-Spizzichinův model opírající se o vlnovou teorii. Podrobnosti z hlediska počítačového vidění lze najít v [37].

#### 8.1.4 Rovnice ozáření a tvar ze stínování

Ozáření  $E(x, y)$  nekonečně malého senzoru světla umístěného v pozici  $x, y$  v obrazové rovině je rovná záři povrchu na odpovídající ploše povrchu určené parametry  $u, v$  popisu plochy v parametrickém tvaru. Předpokládejme, že světelny paprsek není znatelně zeslaben optickým médiem mezi povrchem a senzorem. Tento fakt lze zapsat ve formě důležité rovnice ozáření udávající vztah mezi orientací povrchu a naměřeným jasem v obraze,

$$E(x, y) = \rho(u, v) R(\mathbf{N}(u, v) \cdot \mathbf{L}, \mathbf{N}(u, v) \cdot \mathbf{V}, \mathbf{V} \cdot \mathbf{L}) . \quad (8.15)$$

Proč je rovnice ozáření důležitá? Je to proto, že otevírá cestu, jak na základě změn odrazivosti, projevujících se změnami jasu v sejmutedém obrazu, přímo usuzovat o tvaru povrchu. Takto formulované úloze se říká *tvar ze stínování* (angl. shape from shading). Potíž je, že řešení úlohy je příliš složité, numericky nestabilní a lze je nalézt jen za drastických zjednodušujících předpokladů.

Uvedeme některé omezující podmínky vedoucí ke snížení složitosti úlohy tvaru ze stínování [35]. Předpokládá se, že

- objekt má homogenní odrazivost, tj.  $\rho(u, v)$  je konstanta;
- zdroj světla je velmi vzdálen; potom lze předpokládat konstantní ozáření scény a stejný směr ke zdroji světla pro všechny body povrchu;
- pozorovatel je velmi vzdálen; potom zář vyzařovaná povrchy scény nezávisí na poloze, ale jen na orientaci. Perspektivní zobrazení se zjednoduší na ortografické (tj. rovnoběžné promítání).

Ukažme si rovnici ozáření pro lambertovský povrch, konstantní albedo, jediný vzdálený zdroj světla, vzdáleného pozorovatele ve stejném směru jako je směr ke zdroji světla. Předpokládejme, že funkce odrazivosti  $R$  je vyjádřena v gradientním prostoru  $(p, q)$ ,

$$E(x, y) = \beta R(p(x, y), q(x, y)) . \quad (8.16)$$

$R(p, q)$  udává ozáření v příslušném bodě ve scéně, činitel úměrnosti  $\beta$  pochází z rovnice (8.7) a závisí na  $f$ -čísle objektivu. V tomto případě je zanedbatelná i vinětace objektivu, protože je směr k pozorovateli totožný se směrem ke zdroji světla.

Změřené ozáření  $E$  může být normalizováno a potom lze vynechat činitel  $\beta$ . Potom lze psát rovnici ozáření v nejednodušším tvaru jako

$$E(x, y) = R(p(x, y), q(x, y)) = R\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) . \quad (8.17)$$

Rovnice ozáření tak byla v nejednodušším tvaru vyjádřena jako diferenciální rovnice 1. řádu. Rovnice je obvykle nelineární, protože funkce odrazivosti  $R$  v mnoha případech závisí na gradientu nelieárně. Přesto je rovnice ozáření základním kamenem při snaze získat informaci o orientaci povrchu z intenzitních obrazů.

*Úloha hledání tvaru ze stínování* byla formulována a řešena za značně zjednodušujících předpokladů: předpokládá se lambertovský povrch, ozáření vzdáleným bodovým zdrojem (např. sluncem), je znám součinitel odrazivosti povrchu objektu a orientace normál k povrchu. Cílem je najít normálu  $n$  ke každé elementární ploše povrchu a záři  $L$  povrchu.

Základní potíž při hledání tvaru na základě rozložení jasu (přesněji záře) v obraze vyplývá z toho, že chceme stanovit normálu k povrchu, tedy 2D veličinu, která se v obraze zobrazuje do skalárni veličiny, tj. záře. Platí to i pro nejjednodušší lambertovské objekty. Proto je úloha hledání tvaru ze stínování nedostatečně určena pro všechny sklony povrchu objektu vyjma míst gradientního prostoru  $s$ , kde záře  $L$  vykazuje lokální maxima. Pro všechny ostatní body se získá celý podprostor možných řešení, který odpovídá vrstevnici v gradientním prostoru.

Horn se spolupracovníky [38, 34] navrhli několik numerických postupů pro rekonstrukci tvaru ze stínování. Aby bylo možné řešit nedourčenou úlohu, využili omezení na tvar, které představuje požadavek globální spojitosti povrchu. Horn opírá řešení úlohy tvaru ze stínování o charakteristický proužek, na němž lze pro každý bod povrchu při znalosti orientace zjišťovat orientaci povrchu v lokálním okolí bodu a jedné orientaci. Jelikož každým bodem na povrchu prochází jeden charakteristický proužek, jsou na povrchu singulární body zvané zřídlá a nory (též propasti). V nich charakteristické proužky začínají a končí. Při pohybu z bodu povrchu o známé orientaci lze potom tvar povrchu rekonstruovat. Vhodnými kandidáty pro počátek takové rekonstrukce jsou body na lambertovském povrchu, které odpovídají lokálním maximum záře. V těchto bodech je normála k povrchu rovnoběžná se směrem od zdroje světla, a tudíž známá. Pro rekonstrukci tvaru jsou také důležité body siluety, což jsou body na povrchu, ve kterých povrch z pohledu pozorovatele zakrývá sám sebe. Normálu k povrchu lze v bodech siluety určit jednoznačně. Normála je rovnoběžná s obrazovou rovinou pozorovatele a současně je kolmá na siluetu. I když lze orientaci k povrchu určit jednoznačně, neplatí to o explicitním vyjádření tvaru [67].

Horn upozornil na vlastnosti terminátorů (termín se používá v astronomii) a označuje místa na povrchu, kde v tečné rovině povrchu leží i směr ke zdroji světla. Jinak řečeno, kde zdroj světla zapadá pod horizont. Horn tvrdil, že orientaci povrchu v těchto bodech nelze určit jednoznačně, ale jen s nadbytečností o jednom stupni volnosti. Šára ukázal [83], že pro povrchy s prostorově nezávislou odrazivostí lze určit orientaci jednoznačně. Geometrie zobrazení musí být známa.

Na závěr výkladu o tvaru ze stínování je nutné přiznat, že snaha o rekonstrukci tvaru z jednoho intenzitního obrazu patří stále mezi teoretické práce. Díky silným omezením na povrch (smí být jen lambertovský) a velikou citlivost na šum nejsou dosavadní výsledky prakticky použitelné. Výjimkou je fotometrické stereo.

### 8.1.5 Tvar z fotometrického stereoa

*Fotometrické stereo* je metoda, která umožňuje zjistit orientaci povrchu za předpokladu, že je známá funkce odrazivosti [100]. Nepletěte si fotometrické stereo s obvyklejším geometrickým stereoviděním, které potřebuje obrazy scény alespoň z dvou různých pozorovacích míst.

Uvažujme lambertovský povrch s proměnným albedem  $\rho$ . Klíčovou myšlenkou fotometrického stereoa je představa jednoho pevného pozorovacího místa a kamery při měnícím se osvětlení scény. Za předpokladu, že máme alespoň tři takové snímky lambertovského povrchu, jsme schopni ze změn ve třech obrazových funkcích jednoznačně určit normály k bodům povrchu.

Křivky konstantní odrazivosti na povrchu odpovídají isofotám, tj. křivkám konstantního ozáření  $E$  v obraze. Isofoty jsou v našem speciálním případě lambertovského povrchu polynomy druhého stupně.

Lokální orientace povrchu  $\mathbf{n} = [p, q]$  je v mapě odrazivosti omezena podél těchto křivek druhého stupně. Při změně směru osvětlení se přirozeně nezmění normály k povrchu ani odrazivost, ale mění se mapa odrazivosti  $R(p, q)$ . Tato informace dále omezí přípustné hodnoty normál k povrchu, a to opět polynomem druhého stupně. Dva obrázky a jím odpovídající směry osvětlení ještě nepostačují, aby byla orientace povrchu  $[p, q]$  určena jednoznačně. Proto je potřeba přidat ještě třetí směr a obrázek. Je-li k dispozici ještě více obrázků a směrů, lze řešit přeurovenou soustavu rovnic.

Při praktickém snímání se použije jedna kamera a  $K$  bodových světelných zdrojů,  $K \geq 3$ , se známými intenzitami a směry  $\mathbf{L}_1, \dots, \mathbf{L}_K$ . Při snímání je vždy zapnutý pouze jeden zdroj světla. Snímací pracoviště musí být fotometricky kalibrováno, aby se vzaly v úvahu intenzity jednotlivých zdrojů světla, zesílení kamery a příslušná stejnosměrná složka. Kalibrace je popsána např. v [27]. Po fotometrické kalibraci poskytuje sejmuty obrazy  $K$  odhadů ozáření  $E_i(x, y)$ ,  $i = 1, \dots, K$ .

Jelikož se z povrchu neodrazí veškerá světelná energie, je albedo  $\rho$ ,  $0 \leq \rho \leq 1$ , a projeví se na ozáření obrazu, jak udává rovnice ozáření (8.15). Pro lambertovský povrch se rovnice ozáření zjednoduší na

$$E(x, y) = \rho R(p, q). \quad (8.18)$$

Vzpomeňme na kosinový zákon (8.14), který udává, že mapa odrazivosti pro lambertovský povrch je skalárním součinem mezi normálou k povrchu  $\mathbf{n}$  a směrem ke zdroji světla  $\mathbf{L}_i$ . Po dosazení mapy odrazivosti do rovnice (8.18) obdržíme  $K$  rovnic ozáření

$$E_i(x, y) = \rho \mathbf{L}_i \mathbf{n}, \quad i = 1, \dots, K. \quad (8.19)$$

Pro každý pixel obrazu  $x, y$  získáme vektor s hodnotami ozáření  $\mathbf{E} = [E_1, \dots, E_K]^T$ . Směry ke světelným zdrojům můžeme zapsat ve tvaru matice rozměru  $K \times 3$

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_K \end{bmatrix}. \quad (8.20)$$

Pro každý pixel obrazu můžeme zapsat soustavu rovnic ozáření jako

$$\mathbf{E} = \rho \mathbf{L} \mathbf{n}. \quad (8.21)$$

Matice  $\mathbf{L}$  nezávisí na poloze pixelu v obrazu a můžeme odvodit vektor reprezentující současně albedo i lokální orientaci povrchu. Pro nejjednodušší případ tří světelných zdrojů,  $K = 3$ , můžeme řešení získat inverzí regulární matice  $\mathbf{L}$

$$\rho \mathbf{n} = \mathbf{L}^{-1} \mathbf{E}. \quad (8.22)$$

Jednotkový vektor normály potom můžeme získat jako

$$\mathbf{n} = \frac{\mathbf{L}^{-1} \mathbf{E}}{\|\mathbf{L}^{-1} \mathbf{E}\|}. \quad (8.23)$$

Pro více než tři zdroje světla se použije pseudoinverze obdélníkové matice, a tak se získá řešení ve smyslu nejmenších čtverců

$$\mathbf{n} = \frac{(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{E}}{\|(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{E}\|}. \quad (8.24)$$

Uvědomme si, že se pseudoinverze nebo inverze pro krajní případ  $K = 3$  počítá v každém pixelu obrazu  $x, y$ , a tím se získá odhad odpovídající normály.

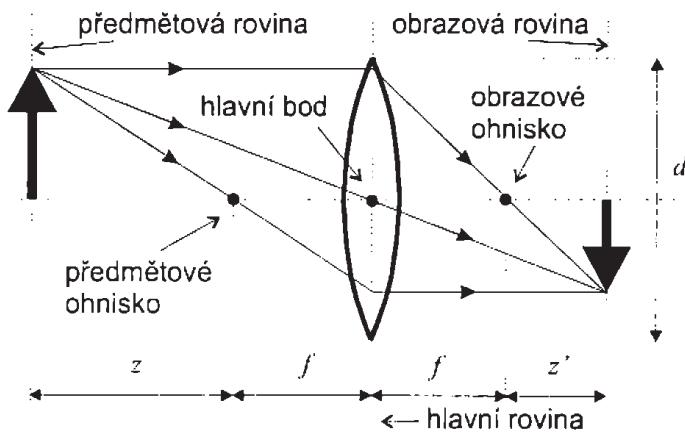
## 8.2 Praktické otázky kolem vzniku intenzitního 2D obrazu

Představme si kameru v jednoduché formě jako zařízení pro snímání intenzitního obrazu. Kamera se skládá z objektivu, snímače obrazu a elektroniky umožňující převést sejmuty obraz k dalšímu zpracování. Objektiv soustřeďuje dopadající energii (fotony) a na snímači se vytváří obraz. Měřenou fyzikální veličinou je ozáření (neformálně jas), jak jsme již vysvětlili v odídle 8.1.2.

### 8.2.1 Optická část

Úkolem objektivu v kamere je promítnout světelné záření na snímač, který převádí ozáření na elektrický signál. Objektiv by měl co nejvěrněji napodobovat ideální projektivní zobrazování (také perspektivní, středové, model dírkové komory), které jsme již znázornili na obr. 6.2 a vyjádřili matematicky vztahem (6.1). Při výkladu se omezíme jen na základní praktické poznatky vyplývající z geometrické optiky. Stranou záměrně ponecháme složitější jevy, jejichž popis vyžaduje komplikovanější matematický model vlnové nebo kvantové optiky.

*Dírková komora* je nereálnou idealizací, protože velmi malou dírkou (též aperturou, aperturní clonou, clonou) neprojde dostatek energie. Další příčinou odchýlení od reality je vlnová povaha světla, která způsobí ohyb (též difrakci), tj. změnu směru šíření záření pro překážky srovnatelné velikosti s vlnovou délkou záření. Nakonec pouze roviná zrcadla přesně vyhovují jednoduchému modelu geometrické optiky zobrazujícímu body na body.



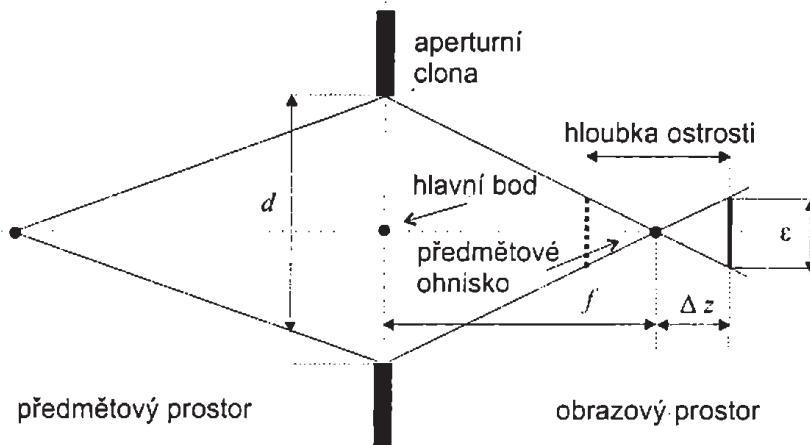
Obrázek 8.5 Tenká čočka.

Ideální model dírkové komory více méně platí pro ideální *tenkou čočku*, která je nejjednodušším objektivem. Projekce paprsku se řídí stejnými zákony jako u dírkové komory a čočka současně sbírá díky konečné apertuře konečné množství světelné energie (fotonů). Paprsek procházející středem čočky (hlavním bodem) se neodklání. U správně zaostřené čočky se procházející paprsky jiných směrů lámou tak, že se protínají s paprskem procházejícím podél optické osy v jednom bodě (*hlavní bod*), viz obr. 8.5. Nedostatkem ideální tenké čočky je skutečnost, že soustřeďuje pouze paprsky vycházející z bodu v rovině kolmé k optické osě ve vzdálenosti  $z$  od středu promítání (hlavního bodu, hlavní roviny), a to podle známé rovnice čočky (odvodíme snadno pomocí podobných trojúhelníků)

v newtonovském tvaru  $\frac{1}{z'} + \frac{1}{f} = \frac{1}{f} + \frac{1}{f+z}$  nebo zapsáno přehledněji  $z \cdot z' = -f^2$ ,

kde  $f$  je ohnisková vzdálenost, tj. vzdálenost mezi hlavním bodem a (obrazovým, předmětovým) ohniskem,  $z'$  je vzdálenost mezi obrazovým ohniskem a obrazovou rovinou,  $z$  je vzdálenost mezi předmětovým ohniskem a předmětovou rovinou. Nechť  $X$  je velikost předmětu (délka šipky v předmětové rovině na obr. 8.5) a  $x$  je velikost obrazu předmětu v obrazové rovině. Zvětšení optické soustavy  $m$  je dáno

$$m = \frac{x}{X} = \frac{f}{z} = \frac{z'}{f}.$$



**Obrázek 8.6** Hloubka ostrosti v obrazovém prostoru optické soustavy.

Posuneme-li obrazovou rovinu mimo obrazové ohnisko, potom se obraz rozostří. Představme si svazek paprsků vycházejících z předmětového ohniska [41], viz obr. 8.6. Bod se zobrazí jako malý kruh o průměru  $\epsilon$ . Nechť  $\Delta z$  je vzdálenost obrazové roviny obrazového ohniska,  $d$  je průměr aperturní clony,  $f$  je ohnisková vzdálenost čočky a  $f + z'$  je vzdálenost hlavního bodu od ohniska. Průměr kruhu  $\epsilon$  se spočte podle podobných trojúhelníků

$$\frac{d}{f + z'} = \frac{\epsilon}{\Delta z}, \quad \text{po úpravě} \quad \epsilon = \frac{d \Delta z}{f + z'}.$$

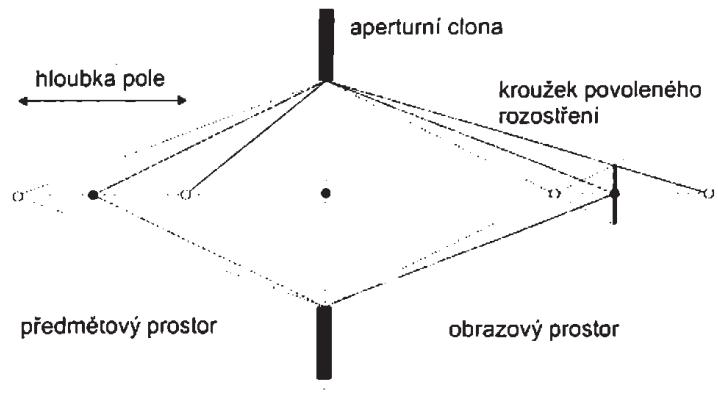
Zavedeme-li  $f$ -číslo objektivu  $n_f = f/d$  (připomeňme, že jsme  $n_f$  již používali v části 8.1.2), můžeme předchozí vztah přepsat na

$$\epsilon = \frac{\Delta z \ f}{n_f(f + z')}.$$

Pojmy hloubka ostrosti (angl. depth of focus) a hloubka pole (angl. depth of field), které nyní zavedeme, vycházejí z představy, že v aplikacích je přípustné, aby obraz byl poněkud rozostřený. Takovou zkušenosť máme z fotografování. Pro digitální obrazy je povolené rozostření úměrné velikosti pixelu. Čím je rozlišení obrazu větší, tím méně může být obraz rozostřen. *Hloubkou ostrosti* je interval povolených posunů obrazové roviny  $(-\Delta z, \Delta z)$ , pro který je průměr kruhu rozostření menší než předdefinovaná hodnota. Hledaná hodnota  $\Delta z$  se vypočte z předchozí rovnice

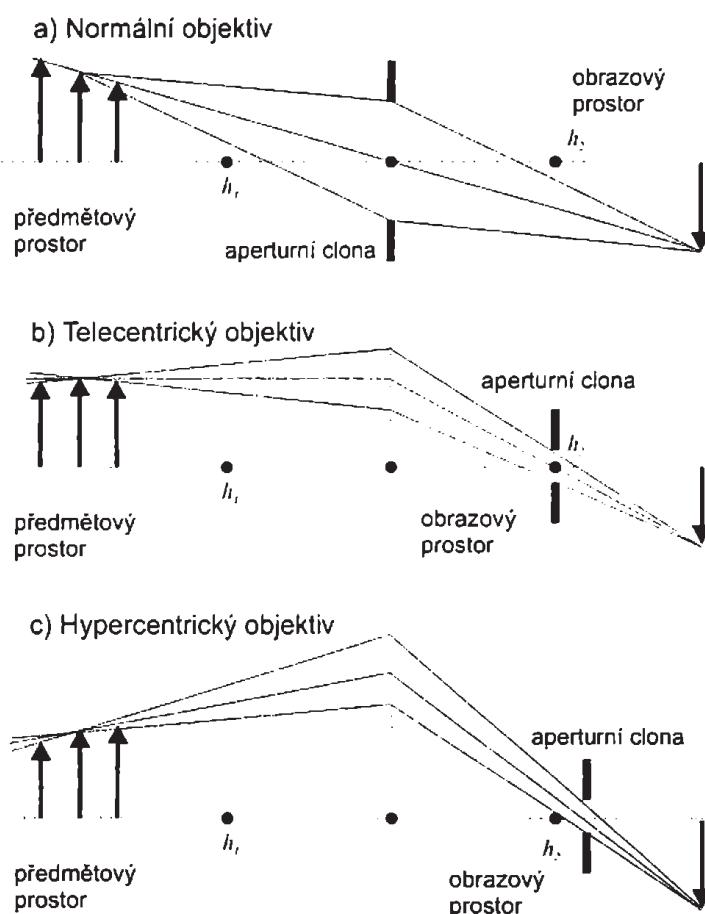
$$\Delta z = n_f \left( 1 + \frac{f}{z'} \right) \epsilon = n_f(1 + m) \epsilon,$$

kde  $m$  je zvětšení optické soustavy. Rovnice ukazuje důležitou roli  $f$ -čísla objektivu pro hloubku ostrosti. Čím je clonová apertura menší, tím je hloubka ostrosti větší.



Obrázek 8.7 Hloubka pole na předmětové straně optické soustavy.

Z hlediska použití optické soustavy je důležitá *hloubka pole* v předmětovém prostoru optické soustavy, která udává rozsah přípustných poloh pozorovaného předmětu při zachovaném požadavku na rozostření nejvýše o  $\varepsilon$ . Pojem je zřejmý z obr. 8.7.



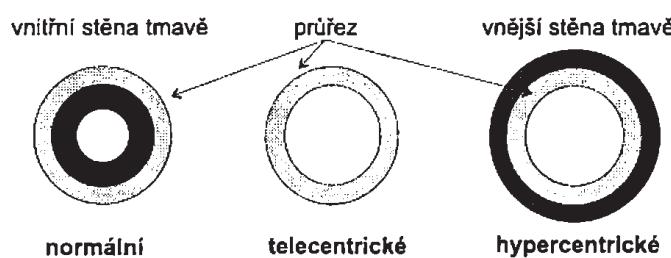
Obrázek 8.8 Normální, telecentrický a hypercentrický objektiv.

Skutečné objektivy složené z několika čoček jsou *tlustými čočkami*, pro něž je možné základní model dírkové komory přirozeně rozšířit. Jedna *hlavní rovina* (tj. rovina kolmá na optickou osu procházející hlavním bodem) se nahradí dvěma hlavními rovinami. U tlusté čočky máme tedy dva hlavní body, předmětový a obrazový. Paprsek vcházející do čočky přes předmětový hlavní bod vychází po průchodu systémem čoček z předmětového hlavního bodu pod

stejným úhlem. Vzdálenost mezi předmětovým a obrazovým hlavním bodem ohniskem udává efektivní tloušťku tlusté čočky. Můžeme zde potom použít jen mírně upravené vztahy pro projekci přes tenkou čočku.

Při optickém měření rozměrů často nelze zaručit, aby měřený předmět byl stále v předmětové rovině. Pokud se objekt k čočce přiblíží, vypadá větší, a když se vzdálí, zdá se být menší. Existuje jednoduchá optická úprava, která potíže při měření rozměru odstraní. Malý aperturní otvor se umístí až do obrazového ohniska, a tak vznikne *telecentrický objektiv*. Tím se dosáhne toho, že obraz vytvářejí pouze paprsky, které jsou přibližně rovnoběžné s optickou osou, což je pro měření rozměrů výhodné. Telecentrický objektiv ilustruje obr. 8.8b. V průmyslových aplikacích získávají telecentrické objektivy na významu. Jejich nevýhodou je, že průměr objektivu musí být přinejmenším tak velký jako měřené rozměry. Díky tomu jsou telecentrické objektivy pro větší průměry (asi  $> 50$  mm) velmi drahé.

Posuneme-li aperturní clonu mezi obrazové ohnisko a obrazovou rovinu, dostaneme hypercentrické zobrazení, u kterého jsou dokonce vzdálenější objekty zobrazovány jako větší, viz obr. 8.8c.



**Obrázek 8.9** Ilustrace zobrazení normálním, telecentrickým a hypercentrickým objektivem pro případ, kdy ze hledí na začátek trubky ve směru její osy.

Znázorněme si ještě zobrazení pomocí normálního, telecentrického a hypercentrického objektivu na příkladě trubky, kterou objektiv pozoruje ve směru její osy, viz obr. 8.9. Normální objektiv díky rozbíhajícím paprskům vidí průřez trubky i vnitřní stěnu a neuvidí vnější stěnu. Telecentrický objektiv pozoruje díky rovnoběžnosti paprsků jen průřez. Hypercentrický objektiv vidí průřez, vnější stěnu a nevidí vnitřní stěnu.

Optické soustavy vykazují řadu dalších vad, díky nímž se projekce odchyluje od popsaného paprskového modelu. Podrobnosti o optických vadách z hlediska analýzy obrazu čtenář najde v [41]. Z bodového předmětu projekcí nevznikne bodový obraz. Výsledkem je distribuce intenzit v obrazu, která se charakterizuje (a) šírkou měřící stupeň rozostření a (b) posunem středu distribuce způsobeným geometrickým zkreslením. Také průchod světla vzduchem či jiným hmotným prostředím způsobuje rozptyl. Obě zmíněné vady lze při hrubém přiblížení modelovat jako pokles vysokých frekvencí ve fourierovském spektru. Při zpracování lze tyto vady kompenzovat filtrem derivačního charakteru.

## 8.2.2 Snímače světelné energie používané v kamerách

Náš výklad v této části bude stručný a pragmatický. Podrobnější vysvětlení principů by vyžadovalo důkladné znalosti fyziky pevných látek. Snímače světelné energie používané v kamerách se dělí do dvou velkých skupin:

*Snímače na fotoemisním principu* využívají vnější fotoelektrický jev. Přijetím fotonu je poskytnut dostatek energie, aby se uvolnil volný elektron. Principu se využívá ve vakuových snímacích elektronkách a fotonásobičích.

*Snímače na fotovoltaickém nebo fotovodivostním principu* se rozšířily s rozvojem polovodičů a využívají vnitřní fotolektrický jev. Energie fotonu způsobí, že elektron opustí svoje valenční pásmo. U fotovodivostních látek se dostane do vodivého pásmá. Množství dopadající světelné energie ovlivňuje makroskopickou vodivost. U fotovoltaických látek je vybuzený elektron zdrojem elektrického napětí, jehož projevem je proud úměrný množství dopadající energie. Příslušnými čidly jsou fotodioda, lavinová fotodioda (zesilovač světla, chováním podobná vakuovému fotonásobiči, zesiluje také šum, používá se např. v kamerách pro noční vidění), fotoodpor a Schottkyho fotodioda (důležitá součást CCD snímačů).

Z praktického hlediska jsou od poloviny sedmdesátých let nejdůležitější CCD snímače (angl. Charge Coupled Devices). V jednotlivých čidlech převádí světelnou energii na elektrický náboj Schottkyho fotodioda. CCD snímač je analogovým posuvným registrem, který sejmuty náboj přesouvá do vyhodnocovací elektroniky. Problémem CCD kamer je vzájemné ovliňování nábojů v sousedních pixelech (angl. blooming), malý rozsah intenzit (2 řády) a z něho vyplývající saturace příslušného pixelu a díky způsobu posouvání náboje nemožnost adresovat jednotlivé pixely. Výroba CCD elementů je poměrně komplikovaná a přece jen méně masová než u polovodičových technologií používaných pro paměti a procesory. Proto cena CCD kamer neklesla pod jednotky tisíc Kč. Výhodou CCD kamer je velká citlivost a poměrně nízký šum.

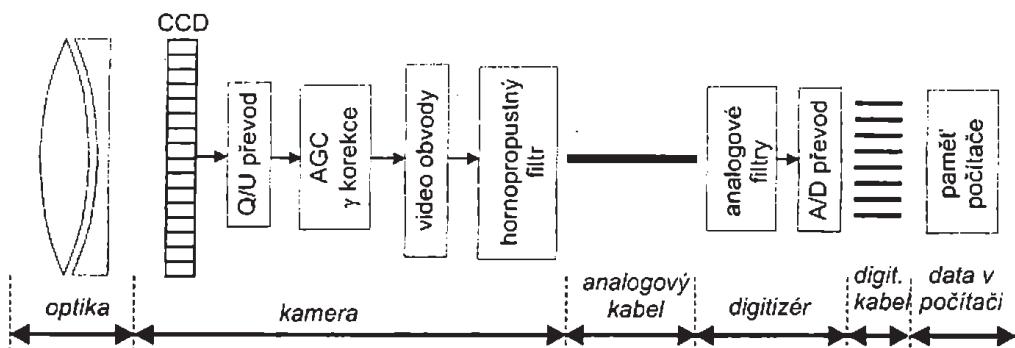
Rozvoj technologie dovolil od poloviny devadesátých let vyrábět snímací čipy také CMOS technologií. Ta je masová, a tudíž levná. Cena čipů se očekává na úrovni jedné nebo několika málo stovek Kč. Jelikož se procesory vyrábějí převážně také CMOS technologií, je možné na jednom čipu nainstalovat kamery i procesor přímo zpracovávající obraz. Výhodou CMOS kamer je větší rozsah intenzit (asi 4 řády), velká rychlosť vyčítání (okolo 100 ns) a náhodný přístup k pixelům. Nevýhodou je asi o řád větší šum než u CCD kamer. CMOS kamery trh teprve dobývají, ale čeká se jejich velké rozšíření zejména pro méně náročné aplikace.

### 8.2.3 Kamera

Kamera se skládá z optické soustavy (objektivu), vlastního snímače a elektroniky, která dovolí sejmuty obraz předat k dalšímu zpracování.

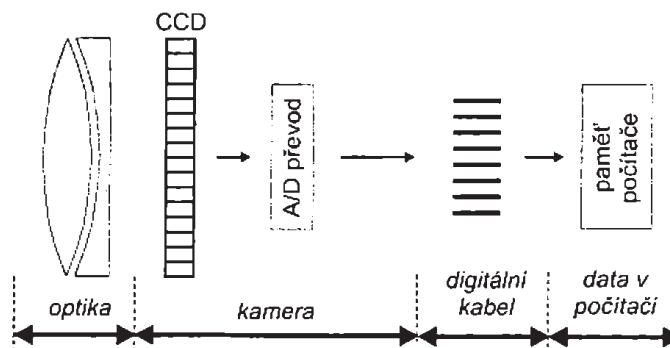
Analogové kamery generují úplný televizní signál, v němž jsou kromě údajů o intenzitě vertikální a horizontální synchronizační pulsy pro postupné vykreslování obrazu po řádcích. Rozklad obrazu může být s prokládanými řádky jako v televizi (50 půlsnímků za sekundu, celý snímek se skládá z 625 řádků) nebo s neprokládanými řádky a vyšší snímkovou frekvencí. U obyčejných analogových CCD kamer nelze příliš spolehnout na časování, což se projevuje tím, že řádky vůči sobě "pluji" (angl. jitter), tj. není zaručeno, že sousedící pixely ve dvou po sobě následujících řádcích na sebe navazují. Pro měřicí účely je tato vlastnost důležitá, a proto existují analogové *pixlově synchronizované kamery*.

Blokové schéma analogové kamery s CCD čipem je na obr. 8.10. Blok AGC (Automatic Gain Control) mění automaticky zesílení kamery, aby se výstup kamery nedostal do saturace (víme už o poměrně malém snímatelném rozsahu intenzit světla), kterému odpovídá napětí  $\langle 0,25V, 1V \rangle$ . Součástí bloku AGC bývá také  $\gamma$  korekce intenzity  $I' = \alpha I^\gamma$ ,  $0 < \gamma \leq 1$ , která se hodí pro zobrazování obrazu člověku a obraz přizpůsobuje logaritmické citlivosti oka na jas. Hodnota  $\gamma = 0,5$  dobře přizpůsobí obraz lidskému pozorovateli. Hodnota  $\gamma < 0,5$  se považuje za silné  $\gamma$  a při hodnotě  $\gamma = 1$  signál prochází beze změny. Upozorněme, že pro měřicí účely je někdy potřebné snímací řetěz radiometricky kalibrovat. V tom případě je nutné mít možnost AGC a  $\gamma$  korekci vypnout. U lepších kamer je to možné zvenku. U levnějších je obvykle možný zásah do elektroniky kamery.



Obrázek 8.10 Analogová CCD kamera.

Videoobvody doplňují informaci o intenzitě synchronizační směsi. Hornopropustný (derivační) filtr zdůrazňuje vysoké frekvence, a tak kompenzuje nežádoucí příspěvek dolnopropustných filtrov způsobujících rozmažání při průchodu signálu vzduchem a v optice. Dále je analogový televizní signál (obvykle koaxiálním kabelem) veden do digitizéru umístěného v počítači. Analogové filtry na vstupu digitizéru hornopropustným filtrem opraví ztrátu vysokých frekvencí v kabelu.



Obrázek 8.11 Digitální CCD kamera.

Blokové schéma digitální kamery je na obr. 8.11. Analogově číslicový převodník získá z intenzity světla číslo, které je dále zpracováno v počítači. Pro připojení se používá buď paralelní nebo sériový způsob využívající protokol IEEE 1394 (fire wire). V prvním případě paralelního připojení je nevýhodou požadavek na krátký kabel (délka okolo 1m).

Srovnajme vlastnosti analogových a digitálních kamer v tabulce.

Analogové kamery	Digitální kamery
+ Levné, rychlé.	- Zatím novinka, dražší, ale cena rychle klesá.
+ Lze dlouhý kabel ( $\approx 10\text{m}$ ).	- Krátký kabel ( $\approx 1,5\text{m}$ ).
- Vícenásobné převzorkování signálu.	+ Jediné převzorkování.
- Šumící, posuv řádků vůči sobě.	+ Bez posunu řádků. Malý šum.
- Nenavazující řádky (angl. jitter).	+ Navazující řádky.

# 9. Geometrické transformace a transformace hodnot jasu

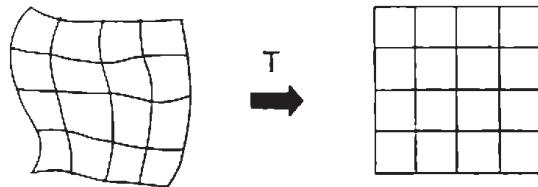
Geometrické transformace 2D obrazu popisují transformaci nosiče obrazové funkce  $f(x, y)$ , tj. souřadnic  $x, y$  při rotaci, zvětšení nebo složitějších zobrazeních. Název transformace hodnot jasu má vystihnout změny hodnoty obrazové funkce  $f(x, y)$ . Místo transformace jasu bychom možná mohli používat fyzikálně správnější název radiometrické transformace, ale upřednostnili jsme intuitivní názornost.

Obě skupiny transformací používané v digitálním zpracování obrazů, tj. geometrické transformace a transformace jasu, probereme postupně. Začneme geometrickými transformacemi.

## 9.1 Geometrické transformace

Geometrické transformace vypočtu na základě souřadnic bodů ve vstupním obrazu souřadnice bodů ve výstupním obrazu. Geometrické transformace jsou obvyklé v počítačové grafice, která často pracuje s objekty ve vektorovém tvaru (reprezentované většinou jako mnohoúhelníky nebo mnohostény) [101]. V digitálním zpracování obrazu navíc geometrické transformace dovolují odstranit geometrické zkreslení vzniklé při pořízení obrazu (např. korekce geometrických vad objektivu kamery, oprava zkreslení družicového snímku způsobená zakřivením zeměkoule). Mezi korekci geometrického zkreslení a geometrickými transformacemi, jako je posunutí nebo rotace, není principiální rozdíl. Použijí se stejné algoritmy. V případě korekci bude mít mluvit o zkresleném, resp. nezkresleném obrazu a u geometrických transformací o novém, resp. původním obrazu. Beze ztráty obecnosti tedy můžeme dále využívat společného aparátu.

Pro většinu úloh digitálního zpracování obrazu stačí, když se omezíme na geometrické transformace v rovině (2D). Zobecnění pro geometrické transformace v prostoru (3D) je přímočaré.



Obrázek 9.1 Ilustrace geometrické transformace v rovině.

Geometrická transformace plošného obrazu je vektorová funkce  $\mathbf{T}$ , která zobrazí bod  $x, y$  do bodu  $x', y'$ . Situaci ilustruje obr. 9.1, kde je bod po bodu transformována část roviny. Transformace  $\mathbf{T}$  je definována dvěma složkovými vztahy

$$x' = T_x(x, y), \quad y' = T_y(x, y). \quad (9.1)$$

Transformační rovnice  $T_x$  a  $T_y$  mohou být buď známé předem, jako je tomu např. v případě rotace, posunu nebo zvětšení obrazu, nebo je možné hledat transformační vztah na základě

znalosti původního i transformovaného obrazu. Při hledání transformace se obvykle využívá několika známých (tzv. vlícovacích) bodů, které v obou obrazech odpovídají identickému objektu a lze je snadno najít. Geometrická transformace sestává ze dvou kroků.

*Transformace souřadnic bodů* najde k bodu ve vstupním obrazu s diskrétními souřadnicemi odpovídající bod ve výstupním obrazu. U výstupního bodu musíme počítat se spojitymi souřadnicemi (reálná čísla), protože poloha výsledného bodu nemusí souhlasit s celočíselnou mřížkou. Tato plošná transformace má tedy bodový charakter.

*Aproximace jasové funkce* hledá celočíselnou hodnotu jasu v celočíselné pozici, která nejlépe odpovídá nově vypočítané neceločíselné poloze  $x', y'$ .

### 9.1.1 Transformace souřadnic bodů

Jak určit souřadnice bodu v obrazu, který je výsledkem geometrické transformace, uvádí v obecném tvaru vztah (9.1). Transformační vztah souřadnic se obvykle approximuje polynomem  $m$ -tého stupně

$$x' = \sum_{r=0}^m \sum_{k=0}^{m-r} a_{rk} x^r y^k, \quad y' = \sum_{r=0}^m \sum_{k=0}^{m-r} b_{rk} x^r y^k. \quad (9.2)$$

Transformace (9.2) je lineární vzhledem ke koeficientům  $a_{rk}, b_{rk}$ . Její výhodou je, že pomocí metody nejmenších čtverců umíme určit koeficienty transformace  $a_{rk}, b_{rk}$  na základě přeuročené množiny dvojic sobě odpovídajících (vlícovacích) bodů ve vstupním a výstupním obrazu  $x, y$  a  $x', y'$ .

V případě, že se žádaná geometrická transformace v závislosti na pozici v obrazu příliš náhle nemění, postačují v rovnici (9.2) approximační polynomy nízkého stupně  $m = 2$  nebo  $m = 3$ .

Rovnice (9.2) se v praxi velmi často nahrazuje *bilineární transformací*, k jejímuž stanovení stačí alespoň čtyři dvojice sobě odpovídajících bodů,

$$\begin{aligned} x' &= a_0 + a_1x + a_2y + a_3xy, \\ y' &= b_0 + b_1x + b_2y + b_3xy. \end{aligned} \quad (9.3)$$

Zvláštním případem bilineární transformace (9.3) je *affinní transformace*, která zahrnuje v praxi potřebnou rotaci, translaci a zkosení

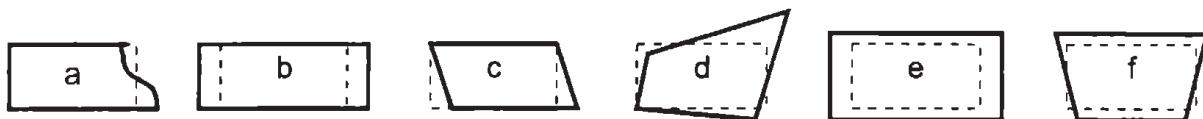
$$\begin{aligned} x' &= a_0 + a_1x + a_2y, \\ y' &= b_0 + b_1x + b_2y. \end{aligned} \quad (9.4)$$

Bodu v rovině odpovídá sloupcový vektor  $[x, y]^T$ . Potom lze affinní transformaci s výhodou vyjádřit pomocí maticových (tj. lineárních) operací.

Zavedeme formalizmus *homogenních souřadnic* (také projektivních souřadnic), jak je obvyklé v teoretické mechanice, projektivní geometrii, počítačové grafice a robotice. To umožní v geometrických transformacích vyjádřit posun, rotaci, affinní transformaci jako součin s jedinou maticí. Základní myšlenkou je reprezentovat bod ve vektorovém prostoru o jednu dimenzi větším. My zde pracujeme s body v rovině. Bod  $[x, y]^T$  se v homogenních souřadnicích vyjádří ve 3D vektorovém prostoru jako  $[\lambda x, \lambda y, \lambda]^T$ , kde  $\lambda \neq 0$ . Pro jednoduchost se obvykle používá jedno z nekonečně mnoha vyjádření bodu v homogenních souřadnicích, a to  $[x, y, 1]^T$ .

Affinní zobrazení (9.4) se po zavedení homogenních souřadnic vyjádří maticově jako

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_0 \\ b_1 & b_2 & b_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (9.5)$$



Obrázek 9.2 Obvyklé typy geometrických zkreslení.

Složitější geometrické transformace (zkreslení) je možné approximovat tak, že se obraz rozdělí na menší obdélníkové podobrazy. Pro každý z podobrazů se použije jednodušší geometrická transformace (např. affinní) odhadnutá z dvojcík sobě odpovídajících bodů. Zkreslení uvnitř každého podobrazu je opravováno zvlášť.

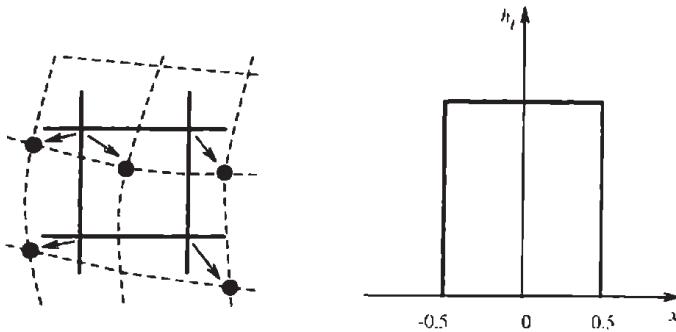
Pro ilustraci uveďme typická geometrická zkreslení, která se vyskytují v dálkovém průzkumu Země. Chybná poloha a orientace senzoru (nebo družice) vzhledem k pozorovanému objektu je hlavním zdrojem chyb rotace, zkosení a nelineárních zkreslení v rádcích. Panoramatické zkreslení (obr. 9.2b) se objevuje u skenerů s rovinným zrcadlem rotujícím konstantní rychlostí. Nelineární zkreslení v rádu (obr. 9.2a) je způsobeno rozdílnou vzdáleností snímaných objektů od zrcadla senzoru. Otáčení zeměkoule při snímání mechanickým skenerem způsobuje chybu zkosení (obr. 9.2c). Obecné lichoběžníkové zkreslení ukazuje obr. 9.2d. Pokud se při snímání změní vzdálenost od pozorovaných objektů, projeví se zkreslení změnou měřítka (obr. 9.2e). Perspektivní zobrazení, které odpovídá nejjednoduššímu modelu dírkové komory approximujícím vlastnosti objektivu kamery ve 3D vidění, je znázorněno na obr. 9.2f.

### 9.1.2 Aproximace jasové funkce

Po geometrické transformaci souřadnic bodů následuje druhý krok, kde se hledá hodnota jasu každého transformovaného bodu. Předpokládejme, že jsme již geometrickou transformací vypočítali souřadnice  $x', y'$  bodů ve výstupním obrazu, a to postupy popsanými v části 9.1.1. Souřadnice  $x, y$  výchozích vzorků obrazové funkce  $f(x, y)$  byly celočíselné, protože byly identické se vzorkovací mřížkou. Po geometrické transformaci (např. otočení o obecný úhel) již souřadnice bodů v obecném případě celočíselné být nemusí. Jelikož také obraz po geometrické transformaci reprezentujeme maticí, máme předepsanou pravoúhlou vzorkovací mřížku i ve výsledném obrazu. Pro názornost si obrazovou funkci představime jako spojitou plochu (např. jako povrch pohoří), kde máme informaci o obrazové funkci (výšce pohoří) jen v bodech vzorkovací mřížky. Po geometrické transformaci nás zajímají hodnoty obrazové funkce (výšky transformovaného pohoří) v místech předepsaných rastrem výstupního obrazu. K dispozici ale máme jen vzorky obrazové funkce v rastru vstupního obrazu (pohoří).

Principiálně správnou představou je *approximovat* (též *interpolovat*) *průběh plochy výstupního obrazu* z dostupných vzorků a stanovit hodnoty v bodech předepsaného rastru. V teorii approximace se obvykle dostupné vzorky proloží polynomem (zde polynomem dvou proměnných  $x, y$ ). Tím získáme analytickou rovnici a po dosazení vypočteme hodnoty transformované obrazové funkce v předepsaném výstupním rastru obrazu. Přesnost interpolace ovlivňuje kvalitu výstupního obrazu. Přesnější postupy pro approximaci používají větší okolí, tj. approximují z více vzorků.

Úloha interpolace jasu se obvykle vyjadřuje způsobem duálním vzhledem k předpisu geometrické transformace (9.1). Approximuje se jas ve vstupním obrazu, který odpovídá jasu hledaného bodu ve výstupní mřížce. Předpokládejme, že chceme vypočítat jas odpovídající pixelu o celočíselných souřadnicích  $(x', y')$  ve výstupní mřížce (na následujících obrázcích bude tato mřížka znázorněna plnou čarou). Souřadnice bodů  $(x, y)$  ve vstupním obrazu lze vypočítat



Obrázek 9.3 Interpolace metodou nejbližšího souseda.

invertováním vztahu (9.1), tj.

$$(x, y) = \mathbf{T}^{-1}(x', y') . \quad (9.6)$$

Jak jsme již zmínili, v obecném případě po výpočtu získáme reálná čísla (v obrázcích obr. 9.3 a 9.4 naznačeno čárkovanými čarami), která nesouhlasí s celočíselným rastrem (znázorněn plnou čarou).

Jedinou dostupnou informací o původně spojité obrazové funkci  $f(x, y)$  je její vzorkovaná verze  $g_s(l \Delta x, k \Delta y)$ . Abychom získali hodnotu jasu v bodě  $(x, y)$ , approximujeme vstupní obraz.

Výsledkem approximace (interpolace) je jas  $f_n(x, y)$ , kde index  $n$  rozlišuje jednotlivé interpolační metody. Jas lze vyjádřit jako dvojrozměrnou konvoluci

$$f_n(x, y) = \sum_{l=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_s(l \Delta x, k \Delta y) h_n(x - l \Delta x, y - k \Delta y) . \quad (9.7)$$

Funkci  $h_n$  se říká *interpolační jádro*. Obvykle se používá interpolační jádro pokrývající jen malé okolí zpracovávaného bodu, aby se ušetřily výpočty. Vně tohoto okolí je hodnota jádra  $h_n$  nulová.

V praxi se pro interpolaci používají jednoduché approximační polynomy. Uvedeme z nich tři od nejjednodušších ke složitějším, a to podle zavedených názvů.

*Metoda nejbližšího souseda* přiřadí bodu  $(x, y)$  hodnotu jasu nejbližšího bodu  $g$  v diskrétní mřížce. V pravé části obr. 9.3 je znázorněno interpolační jádro  $h_1$  v jednorozměrném případě. Levá strana obr. 9.3 přibližuje, jak je přiřazena nová hodnota jasu. Čárkované čáry ukazují, jak je celočíselný rastr výstupního obrazu zobrazen inverzní rovinou transformací (9.6). Plné čáry ukazují vzorkovací mřížku vstupního obrazu.

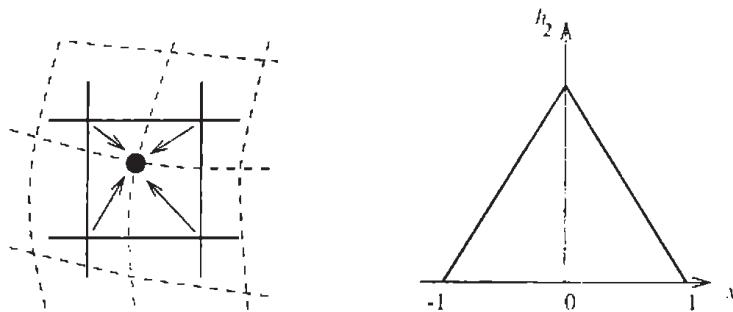
Interpolace metodou nejbližšího souseda je dána

$$f_1(x, y) = g_s(\text{round}(x), \text{round}(y)) . \quad (9.8)$$

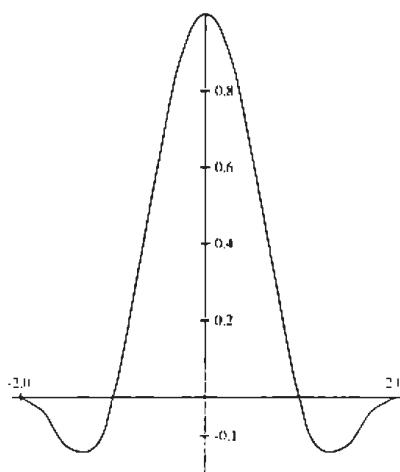
Chyba v poloze bodu po interpolaci metodou nejbližšího souseda je nejvýše půl pixelu. Tato chyba je v obrázcích dobře zřetelná u objektů s přímočarými obrysami natočenými šikmo vůči rastru, které jsou schodovité.

*Lineární interpolace* využije okolí čtyř bodů sousedících se zpracovávaným bodem  $(x, y)$  a předpokládá, že obrazová funkce je lineární kombinací jasu těchto čtyř bodů. Vliv každého ze čtyř bodů v lineární kombinaci je úměrný jeho blízkosti ke zpracovávanému bodu.

Lineární interpolace je znázorněna na obr. 9.4. Levá část obr. ukazuje, jak se čtyři body v okolí podílejí na interpolaci, a pravá část demonstriuje interpolační jádro v jednoroz-



Obrázek 9.4 Lineární interpolace.



Obrázek 9.5 Interpolaci jádro pro bikubickou interpolaci.

měrném případě. Lineární interpolace je dána vztahem

$$\begin{aligned} f_2(x, y) = & (1-a)(1-b)g_s(l, k) + a(1-b)g_s(l+1, k) \\ & + b(1-a)g_s(l, k+1) + abg_s(l+1, k+1), \end{aligned} \quad (9.9)$$

kde  $l = \text{round}(x)$ ,  $a = x - l$ ,  
 $k = \text{round}(y)$ ,  $b = y - k$ .

Lincární interpolace způsobuje malý pokles rozlišení, protože obraz mírně rozinazává díky tomu, že hodnoty obrazu jsou vyhlazovány lineárním filtrem. Rozmazání je ovšem menší zlo než schodovitost u interpolace metodou nejbližšího souseda. Zlepšení je nejvíce patrné na šíkmých čarách.

*Bikubická interpolace* dále zpřesňuje model obrazové funkce tím, že je lokálně interpolován bikubickým polynomem. Pro výpočet se používá okolí složené z 16 bodů.

Jelikož by dvojrozměrné znázornění bylo nepřehledné, ukážeme interpolaci jádro v jednorozměrném případě

$$h_3 = \begin{cases} 1 - 2|x|^2 + |x|^3 & \text{pro } 0 \leq |x| < 1, \\ 4 - 8|x| + 5|x|^2 - |x|^3 & \text{pro } 1 \leq |x| < 2, \\ 0 & \text{jinde.} \end{cases} \quad (9.10)$$

Interpolaci jádro je vidět na obr. 9.5. Jeho tvar by ve 2D připomínal mexický klobouk. Bikubická interpolace zachovává mnohem lépe detaily v obraze než dvě výše zmíněné metody. Proto se používá v rastrových displejích a zobrazovacích programech, které dovolují zoom v libovolném bodě.

## 9.2 Transformace hodnot jasu

Transformace hodnot jasu lze rozdělit do dvou skupin: jasové korekce a modifikace jasové stupnice. U jasových korekcí závisí jas v bodě výstupního obrazu na jasu odpovídajícího bodu ve vstupním obrazu (a případně na jasu jeho malého okolí). U modifikací jasové stupnice je jen určitá hodnota jasu ve vstupním obrazu transformována na jinou výstupní hodnotu, a to bez ohledu na polohu v obrazu.

### 9.2.1 Jasové korekce

První skupinou metod jsou jasové korekce. Snímací a digitalizační zařízení má mít v ideálním případě stejnou citlivost bez ohledu na umístění bodu v obrazu. V praxi však není tento požadavek v řadě případů splněn. V optických soustavách je světlo procházející dále od optické osy více zeslabováno (v optice se této vadě objektivu říká vinětace). Také světlocitlivý element kamery, např. CCD prvek, nemusí být stejně citlivý ve všech svých bodech. Zdrojem poruch je i nerovnoměrné osvětlení snímané scény. Zákeřné jsou také prachové částice na krycím sklu optického senzoru kamery. Částečky nejsou zobrazeny ostře, jsou okem těžko pozorovatelné, ale přesto pohlcují světlo. Prach tak způsobuje zdánlivý pokles citlivosti senzoru. V kontrastních scénách s mnoha objekty je změna citlivosti těžko patrná. Mnohem zřejmější se stane při pohledu na scénu o stálém jasu.

Jsou-li uvedené poruchy systematické (míněno deterministické, jako protiklad k náhodné, o odstranění náhodných vlivů se zmíníme v sekci 11.1), lze je potlačit jasovými korekcmi, a to na základě znalosti odchylky citlivosti každého bodu obrazu od ideální převodní charakteristiky. Nejčastěji se předpokládá jednoduchý model porušení obrazu, a to multiplikativním koeficientem  $e(x, y)$ . Pro každý bod  $x, y$  původního obrazu  $g(x, y)$  získáme na výstupu hodnotu jasu zkresleného obrazu  $f(x, y)$  podle

$$f(x, y) = e(x, y) g(x, y). \quad (9.11)$$

Při systematické degradaci  $e$  stačí při stálých snímacích podmírkách sejmout obraz o známém průběhu jasové funkce  $g(x, y)$ . Nechť je pro jednoduchost etalonem obraz o konstantním jasu  $c$  (etalonová šedá plocha), který po sejmnutí a digitalizaci označíme  $f_c(x, y)$ . Potom můžeme systematické chyby ve snímacím řetězu korigovat podle vztahu

$$g(x, y) = \frac{f(x, y)}{e(x, y)} = \frac{c f(x, y)}{f_c(x, y)}.$$

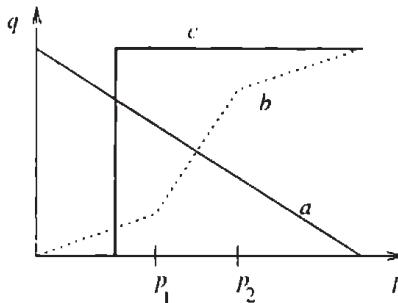
Popsaná korekce chyb etalonovou šedou plochou přirozeně platí pro stálé snímací podmínky. Jinak je třeba korekční transformační matici stanovit po každém kalibrování snímacího řetězce.

U digitálních obrazů jsou hodnoty jasu omezeny v určitém intervalu. Výpočet korigovaných hodnot podle vztahu (9.11) však může dát výsledky i mimo přístupný interval jasů. Jedním řešením je nahrazena hodnot jasu vně intervalu jeho krajními hodnotami. Druhou možností je zhuštění nebo posun jasové stupnice tak, aby odpovídala přípustným hodnotám.

### 9.2.2 Transformace jasové stupnice

Transformace jasové stupnice nezávisí na poloze v obrazu, je tedy stejná pro všechny pixely obrazu. Transformace  $\mathcal{T}$  výchozí stupnice jasu  $p = \langle p_0, p_k \rangle$  na novou stupnici  $q = \langle q_0, q_k \rangle$  je dána vztahem

$$q = \mathcal{T}(p). \quad (9.12)$$



**Obrázek 9.6** Obvyklé transformace jasové stupnice; (a) negativ; (b) zvětšení kontrastu mezi jasy  $p_1$  a  $p_2$ ; (c) prahování, jehož výsledkem je obrázek obsahující jen černou a bílou (každý pixel lze kódovat jediným bitem).

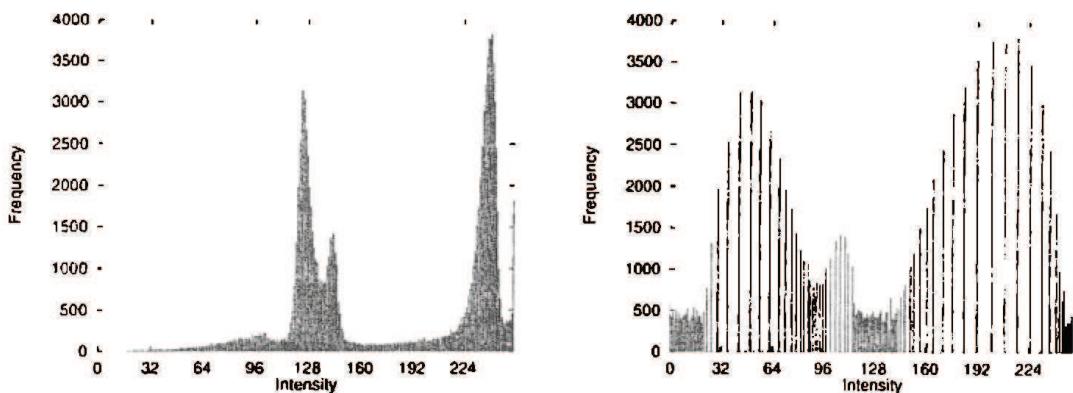
Příklady obvyklých transformací jasové stupnice jsou na obr. 9.6. V displejích nebo kamerech se často používá také logaritmická stupnice, protože odpovídá logaritmické odezvě detektorů lidského oka citlivých na jas. Díky tomu můžeme okem pozorovat jasy, jejichž intenzita se liší až o 9 řádů. Tato logaritmická transformace se v prodávaných přístrojích označuje názvem  $\gamma$  (gama) korekce. Má-li být kamera použita pro měřicí účely, bývá potřebné  $\gamma$  korekci vypnout. U lepších kamer to bývá možné.

Z praktického hlediska je transformace jasové stupnice důležitá zejména pro úpravy obrazu, které zajížfují pozorovateli snazší interpretaci vizualizovaného obrazu. Příkladem může být snaha o zvýšení kontrastu původně nekontrastního rentgenového obrazu. Upozorněme, že pro automatickou analýzu, kde obrazová data člověk neinterpretuje, nemají transformace jasové stupnice žádný význam. Mohou dokonce uškodit, protože mohou vést ke ztrátě informace, kdyby transformace jasové stupnice nebyla invertovatelná (tj. nebyla by prostým zobrazením).

Digitální monochromatické obrazy mají typicky velmi omezený počet jasových úrovní, obvykle 256 úrovní reprezentovaných v osmi bitech. Transformace jasové stupnice se poměrně snadno technicky realizují. V obrazových periferiích je vytvořena vyhledávací tabulka (angl. table look up) pomocí paměti o tolika paměťových místech, kolik je jasových úrovní. Nové hodnoty jasu, které jsou výsledkem transformace, jsou obsahem paměti. Vyhledávací tabulka bývá zařazena do cesty obrazového signálu, a tudíž bývá možné pozorovat výsledek transformace v reálném čase.

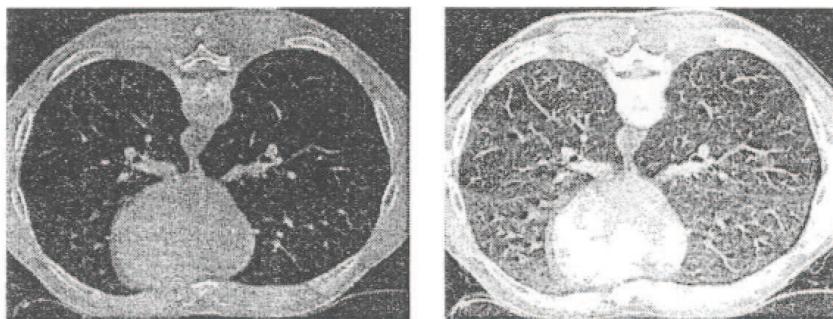
Podobný princip lze použít i pro barevné displeje. Vyhledávací tabulka je zařazena v každé ze tří barevných složek R, G, B. Hodnoty obrazové funkce  $f(x, y)$  jsou potom interpretovány jako indexy do vyhledávací tabulky, a proto zobrazeny barevně. V terminologii osobních počítačů se takovým vyhledávacím tabulkám říká paleta. Popsaný princip má dvě hlavní použití:  
**Pseudobarevné obrazy.** Výchozí obraz je monochromatický. Přes vyhledávací tabulku je zobrazen pseudobarevně, což znamená, že jasům jsou přiřazeny barvy. Důvodem je skutečnost, že lidské oko je více citlivé na změnu barvy než na změnu jasu, a člověk tak rozliší více "jasových" detailů.

**Paletové obrazy.** Výchozí obraz byl barevný. Aby se uspořila paměť pro jeho uchování, byl ztrátově (tj. po dekomprezji nezískáme totéž, ale obraz zachovávající podstatné rysy) komprimován do jediného obrazu s  $M$  jasovými úrovněmi ( $M$  je mocnina 2, typicky  $M = 256$ ). Celý barevný prostor obrazu se rozdělí do  $M$  souvislých shluků, které reprezentují podobné barvy. Pro každý shluk se zvolí jeden typický zástupce, ležící "uprostřed" shluku. Tři vyhledávací tabulky jsou vytvořeny tak, aby  $M$  hodnot obrazové funkce představovaly indexy do tří vyhledávacích tabulek a přes ně zobrazovaly typické barevné zástupce. Takovým obrazům se říká paletové obrazky. Obvykle se s obrazem uchovává i paleta, tj. tříkrát  $\log_2 M$  bitů.



**Obrázek 9.7** Ekvalizace histogramu: Vlevo histogram původního obrazu. Vpravo histogram výsledného kontrastnějšího obrazu, viz obr. 9.8.

Pro optimální zvýšení kontrastu monochromatického obrazu se velmi často používá ekvalizace (vyrovnaní) histogramu. Ve vyrovnaném histogramu obrazu po transformaci jasové stupnice jsou jednotlivé jasové úrovně zastoupeny zhruba stejně četně. Ideální by byl histogram, kde by byly všechny četnosti stejné, ale ve světě digitálních obrázků s poměrně málo jasovými úrovněmi (stovky) a mnoha pixely (statisice) to většinou nejde. Ekvalizace zvýší kontrast pro úrovně jasu blízko maxim histogramu a sníží kontrast blízko minim histogramu. Příklad histogramu před a po ekvalizaci je uveden na obr. 9.7.



**Obrázek 9.8** Ekvalizace histogramu: Vlevo původní obraz (tomografický řez plízemi). Vpravo zvýšení kontrastu po ekvalizaci.

Na obr. 9.8 je uvedena dvojice intenzitních obrazů (jeden řez z rentgenového tomografu napříč hrudníkem zobrazující plíce), jimž odpovídají histogramy z obr. 9.7. Je vidět, že u obrazu vpravo po ekvalizaci se zvýšil kontrast.

Odvodíme postup ekvalizace histogramu. Označíme histogram výchozího obrazu  $H(p)$ , jehož jasová stupnice je  $p = \{p_0, p_k\}$ . Cílem je najít monotónní transformaci jasové stupnice  $q = \mathcal{T}(p)$ , aby výsledný histogram  $G(q)$  byl rovnoměrný pro celý výstupní interval jasů  $q = \{q_0, q_k\}$ .

Histogram můžeme chápat jako kvantovanou diskrétní hustotu pravděpodobnosti. Z požadavku na monotónnost zobrazení  $\mathcal{T}$  plyne

$$\sum_{i=0}^k G(q_i) = \sum_{i=0}^k H(p_i). \quad (9.13)$$

Sumy v rovnici (9.13) lze chápat jako diskrétní distribuční funkce. Předpokládejme, že obraz má  $N$  řádků a  $N$  sloupců. Ekvalizovaný histogram  $G(q)$  odpovídá rovnoměrnému rozdělení  $f$ ,

jehož hustota pravděpodobnosti je konstantní.

$$f = \frac{N^2}{q_k - q_0} . \quad (9.14)$$

Výsledek z rovnice (9.14) dosadíme za levou stranu rovnice (9.13). Ideálně ekvalizovaný histogram lze získat pouze v abstraktním případě, kdy by bylo rozdělení (zde odpovídající histogramu) spojité. V odvození tuto abstrakci pro jednoduchost použijeme. V tomto případě z rovnice (9.13) získáme

$$N^2 \int_{q_0}^q \frac{1}{q_k - q_0} ds = \frac{N^2(q - q_0)}{q_k - q_0} = \int_{p_0}^p H(s) ds . \quad (9.15)$$

Hledaná transformace jasové stupnice je

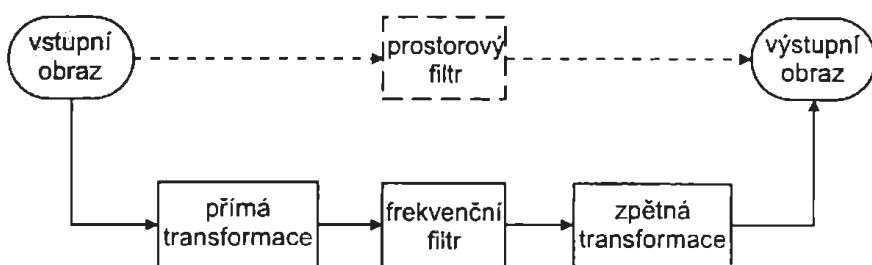
$$q = \mathcal{T}(p) = \frac{q_k - q_0}{N^2} \int_{p_0}^p H(s) ds + q_0 . \quad (9.16)$$

Integrálu v rovnici (9.16) se říká *kumulativní histogram*. Pro digitální obrazy je approximován sumou, a proto nebývá histogram zcela vyrovnaný, jak jsme již viděli na obr. 9.7. Diskrétní approximace transformace jasové stupnice (9.16) je

$$q = \mathcal{T}(p) = \frac{q_k - q_0}{N^2} \sum_{i=p_0}^p H(i) + q_0 . \quad (9.17)$$

# 10. Diskrétní lineární integrální transformace

Zpracování obrazů, které využívá spojité nebo diskrétní lineární integrální transformace, jako např. Fourierovu, lze považovat za klasické techniky. Jejich základní postupy se bouřlivě rozvíjely v 60. letech a je z nich patrná jejich inspirace v lineární teorii signálů, kterou čtenář již zná z první části tohoto skripta. Jejich aplikační kolébkou byly teorie elektrických obvodů, radiotechnika a teorie dynamického řízení.



Obrázek 10.1 Obraz lze lineárními filtry zpracovat buď v prostorové, nebo frekvenční oblasti.

V této kapitole se budeme opírat o představu filtru jako bloku zpracování (viz obr. 10.1), na jehož vstupu i výstupu je dvojrozměrný signál (obraz). Jelikož nyní budeme mít na mysli lineární filtry, budeme se moci vydat dvěma ekvivalentními cestami:

*Filtrace v prostorové oblasti* (pro jednorozměrné signály by se řeklo v časové oblasti) obraz zpracovává jako lineární kombinaci vstupního obrazu s koeficienty filtru (často s mnohem menším definičním oborem než obraz, tzv. lokální filtry). Základním matematickým nástrojem bude konvoluce. Jde o rozšíření konvoluce z první části skripta na dvojrozměrné signály. Tento přístup je na obr. 10.1 označen čárkovaně a bude podrobněji rozebrán v kapitole 11.

*Filtrace ve frekvenční oblasti* nejdříve převede obraz lineární integrální transformací do "frekvenční reprezentace", kde se filtruje a výsledek filtrace se inverzní lineární integrální transformací převede opět na obraz. V mnoha případech je vyjádření filtru ve frekvenční reprezentaci názornější. Uvidíme, že možných transformací je více. Pro začátek stačí, když si čtenář představí Fourierovu transformaci. Na obr. 10.1 je cesta přes frekvenční reprezentaci naznačena plnou čarou. Budeme se jí zabývat v této kapitole, přičemž navážeme na znalosti o Fourierově transformaci z části skripta o teorii signálů.

## 10.1 Základní teorie

Nechť je obraz reprezentován maticí celých čísel  $f$ . Obrazová transformace obvykle zpracuje celý obraz nebo jeho část (obvykle obdélníkovou). Předpokládejme, že zpracovávaný obraz má

rozměry  $M \times N$ . Obraz zapíšeme pomocí

$$\mathbf{f} = \begin{bmatrix} f(0,0) & f(0,1) & f(0,N-1) \\ \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) & f(M-1,N-1) \end{bmatrix}. \quad (10.1)$$

Transformační matice  $\mathbf{P}$  a  $\mathbf{Q}$  rozměru  $M \times M$ , resp.  $N \times N$  se použijí k převedení matice  $\mathbf{f}$  na matici  $\mathbf{F}$ . Jak  $\mathbf{f}$  tak i  $\mathbf{F}$  jsou rozměru  $M \times N$ . Převod je podle vztahu

$$\mathbf{F} = \mathbf{P} \mathbf{f} \mathbf{Q}, \quad (10.2)$$

který lze vyjádřit i pomocí sum jako

$$F(u, v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} P(u, m) f(m, n) Q(n, v), \quad u = 0, 1, \dots, M-1; v = 0, 1, \dots, N-1. \quad (10.3)$$

Pokud jsou matice  $\mathbf{P}$  a  $\mathbf{Q}$  regulární (tj. mají nenulové determinanty), existují inverzní matice  $\mathbf{P}^{-1}$  and  $\mathbf{Q}^{-1}$ . Inverzní transformaci lze zapsat

$$\mathbf{f} = \mathbf{P}^{-1} \mathbf{F} \mathbf{Q}^{-1}. \quad (10.4)$$

Připomeneme si několik pojmu a faktů z teorie matic, které budeme potřebovat. Označením  $\mathbf{M}^T$  vyjádříme transponovanou matici k matici  $\mathbf{M}$ :

- $\mathbf{M}$  je *symetrická*, když  $\mathbf{M} = \mathbf{M}^T$ .
- $\mathbf{M}$  je *ortogonální*, když  $\mathbf{M}^T \mathbf{M} = \mathbf{I}$ , kde  $\mathbf{I}$  označuje jednotkovou matici.
- Pro každou symetrickou, reálnou a ortogonální matici platí:  $\mathbf{M}^{-1} = \mathbf{M}$ .
- Komplexní matice  $\mathbf{C}$  je *hermitovská*, když  $\mathbf{C}^{*T} = \mathbf{C}$ , kde matice  $\mathbf{C}^*$  vznikne z matice  $\mathbf{C}$  tak, že každý její prvek je nahrazen komplexně sdruženým prvkem.
- Komplexní matice  $\mathbf{C}$  je *unitární*, když  $\mathbf{C}^{*T} \mathbf{C} = \mathbf{I}$ .
- Pro každou čtvercovou, komplexní, hermitovskou a unitární matici platí:  $\mathbf{C}^{-1} = \mathbf{C}$ .

Pokud jsou jak  $\mathbf{P}$  tak i  $\mathbf{Q}$  symetrické, reálné a ortogonální matice, potom platí

$$\mathbf{F} = \mathbf{P} \mathbf{f} \mathbf{Q}, \quad \mathbf{f} = \mathbf{P} \mathbf{F} \mathbf{Q}. \quad (10.5)$$

Příslušná transformace je *ortogonální*. Jsou-li matice  $\mathbf{P}$ ,  $\mathbf{Q}$  komplexní, pak za předpokladu, že jsou matice také hermitovské a unitární, stále platí rovnice (10.5).

## 10.2 Fourierova transformace

Jednorozměrná Fourierova transformace byla probrána v první části skripta o jednorozměrných signálech. Vlastnosti jako linearita, invariantnost vůči posuvu, obraz konvoluce, vztah ke korelace, atd. jsou zachovány i ve dvojrozměrném případě. Fourierovu transformaci zapíšeme v této kapitole zavedeném maticovém formalismu. Nechť  $\Phi_{JJ}$  je transformační matice rozměru  $J \times J$  a  $j$  je imaginární jednotka:

$$\Phi_{JJ}(k, l) = \frac{1}{J} \exp\left(-j \frac{2\pi}{J} kl\right), \quad k, l = 0, 1, \dots, J-1. \quad (10.6)$$

Diskrétní Fourierovu transformaci (DFT) definuje podle vztahu (10.2), když přiřadíme  $\mathbf{P} = \Phi_{MM}$ ,  $\mathbf{Q} = \Phi_{NN}$ , a to v maticovém zápisu

$$\mathbf{F} = \Phi_{MM} \mathbf{f} \Phi_{NN}. \quad (10.7)$$

nebo zápisu se sumami

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \exp \left[ -2\pi j \left( \frac{mu}{M} + \frac{nv}{N} \right) \right], \quad (10.8)$$

$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Inverzní transformační matice  $\Phi_{JJ}^{-1}$  je

$$\Phi_{JJ}^{-1}(k, l) = \exp \left( \frac{2\pi j}{J} kl \right) \quad (10.9)$$

a inverzní Fourierovu transformaci můžeme zapsat jako

$$f(m, n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \exp \left[ 2\pi j \left( \frac{mu}{M} + \frac{nv}{N} \right) \right], \quad (10.10)$$

S ohledem na rychlou implementaci si všimněme, že rovnici (10.8) lze přepsat jako

$$F(u, v) = \frac{1}{M} \sum_{m=0}^{M-1} \left[ \frac{1}{N} \sum_{n=0}^{N-1} \exp \left( \frac{-2\pi j nv}{N} \right) f(m, n) \right] \exp \left( \frac{-2\pi j mu}{M} \right), \quad (10.11)$$

$$u = 0, 1, \dots, M-1; \quad v = 0, 1, \dots, N-1.$$

Výraz v hranatých závorkách odpovídá jednorozměrné DFT  $m$ -té řádky obrazu a lze ho vypočítat standardní rychlou Fourierovou transformací (FFT, za předpokladu  $N = 2^k$ ). Každý řádek je nahrazen svou DFT a potom je vypočtena jednorozměrná DFT každého sloupce.

Důležitou vlastností DFT je její periodicitu. Transformaci jsme definovali jako matici s prvky  $F(u, v)$  pro  $u = 0, 1, \dots, M-1; v = 0, 1, \dots, N-1$ . Pokud jsou povoleny i jiné hodnoty  $u, v$ , potom lze zavést periodickou transformaci  $F$  (viz (10.8)). Periodický obraz  $f$  je definován jako

$$\begin{aligned} F(u, -v) &= F(u, N-v), & f(-m, n) &= f(M-m, n), \\ F(-u, v) &= F(M-u, v), & f(m, -n) &= f(m, N-n). \end{aligned} \quad (10.12)$$

Pro celá čísla  $a, b$  také platí

$$F(aM+u, bN+v) = F(u, v), \quad f(aM+m, bN+n) = f(m, n). \quad (10.13)$$

Když si vzpomeneme na rovnici (10.10), vidíme, že můžeme obrazové funkce  $f(m, n)$  interpretovat jako lineární kombinaci periodických vzorů  $\exp[2\pi j((mu/M) + (nv/N))]$  a výsledek DFT, tedy  $F(u, v)$ , může být chápán jako váhová funkce. Výsledkem Fourierovy transformace reálné funkce je komplexní funkce

$$F(u, v) = R(u, v) + jI(u, v), \quad (10.14)$$

kde  $R(u, v)$ , resp.  $I(u, v)$  jsou reálnou a imaginární složkou  $F(u, v)$ . Modul (velikost)  $|F(u, v)|$  se nazývá amplitudové frekvenční spektrum obrazu  $f(m, n)$ . Dále se používají pojmy: fázové spektrum  $\phi(u, v)$  a výkonové spektrum  $P(u, v)$ . Amplitudové frekvenční spektrum je definováno vztahem

$$|F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)}, \quad (10.15)$$

a fázové spektrum jako

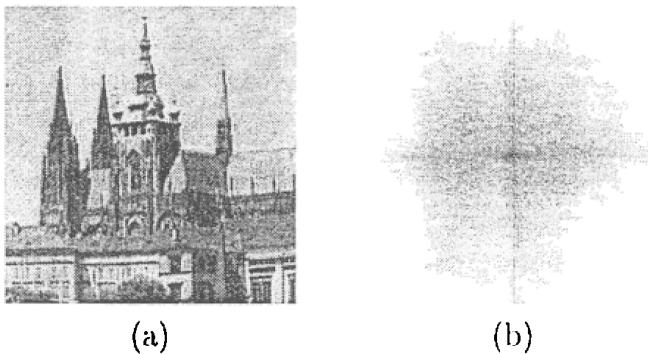
$$\phi(u, v) = \tan^{-1} \left[ \frac{I(u, v)}{R(u, v)} \right]. \quad (10.16)$$

Poznamenejme, že je potřeba dát pozor na znaménka kvadrantů, protože funkce tan je periodická s periodou  $\pi$  a  $\phi$  probíhá od 0 do  $2\pi$ .

*Výkonové spektrum* (nazývané také výkonová spektrální hustota) je definováno pomocí

$$P(u, v) = |F(u, v)|^2 = R^2(u, v) + I^2(u, v). \quad (10.17)$$

Příklad intenzitního obrazu (v prostorové oblasti) a jeho frekvenčního spektra (ve frekvenční oblasti) je na obr. 10.2. Všimněte si, že ve spektru je patrný kříž se středem ve středu obrázku. Diskrétní Fourierova transformace předpokládá, že obraz je jedna perioda 2D periodické funkce. Přepokládá tedy, že okraje obrazu budou stejné a budou na sebe navazovat. Jestliže tento předpoklad není splněn, jako v případě obrazu Pražského hradu, jsou ve fourierovském obrazu potřebné vysoké frekvence, aby bylo možné rekonstruovat nespojitosti na krajích obrazu. Ve spektru na obr. 10.2 odpovídá nespojitostem centrální kříž. Jde o analogii případu obdélníkového průběhu, pro jehož rekonstrukci bychom potřebovali nekonečně mnoho harmonických složek. Při pečlivějším pohledu na spektrum si lze všimnout, že kříže jsou dva. Výraznější kříž se kryje s osami  $u, v$  spektra a vznikl transformací okrajů intenzitního obrazu. Druhý kříž je natočen asi o  $10^\circ$  proti směru hodinových ručiček vůči prvnímu kříži a nedosahuje až do krajů obrazu. Pochází z převažujících směrů jasových rozhraní v obrázku (gradientu obrazové funkce). Poznamenejme, že hranice objektů vnímaných v obrazu člověkem jsou pootočeny proti směru jasových rozhraní (ve frekvenčním spektru jde o harmonické bázové funkce). Proto svislé směry kříže ve spektru odpovídají vodorovným hranám v intenzitním obrázku a naopak.



**Obrázek 10.2** Fourierovo frekvenční spektrum: (a) Vstupní obraz, (b) spektrum jako intenzitní obraz – tmavé pixely odpovídají vysokým spektrálním hodnotám.

Významné je použití Fourierovy transformace pro výpočet konvoluce dvou signálů (obrazů). Vzpomeňme, že konvoluce popisuje průchod obrazu lineárním filtrem, což je základ filtrace obrazu v prostorové oblasti, viz obr. 10.1. Vypočteme-li konvoluci dvou (periodických) obrazů  $f, h$  o stejných rozměrech  $M \times N$ , získáme (periodický) obraz  $g$

$$g(a, b) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)h(a-m, b-n). \quad (10.18)$$

Výpočty v indexech matice  $h(a-m, b-n)$  musí být počítány modulo  $M$ , resp.  $N$ . *Konvoluční věta* říká, že  $f, g, h$  a jejich Fourierovy obrazy  $F, G, H$  jsou vázány rovnicí

$$G(u, v) = F(u, v)H(u, v), \quad (10.19)$$

která vyjadřuje součin po prvcích. Můžeme tedy psát

$$g(a, b) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) H(u, v) \exp \left[ 2\pi j \left( \frac{au}{M} + \frac{bv}{N} \right) \right]. \quad (10.20)$$

Využitím tohoto vztahu lze podstatně snížit výpočetní nároky, což se výrazněji projeví pro konvoluce s konvolučními jádry (v části skripta o signálech byly nazývány impulsními odezvami) většími než  $7 \times 7$ .

V této sekci jsme si všimli, že Fourierova transformace může být chápána jako lineární kombinace ortogonálních harmonických průběhů. Koeficienty Fourierovy transformace slouží jako koeficienty v lineární kombinaci. Harmonické složky jsou uspořádány podle frekvence. Čím vyšší frekvence složkového signálu, tím detailnější příspěvek k popisu signálu tato složka poskytuje. V následujících sekcích ukážeme, jak využít myšlenku rozkladu libovolného signálu (obrazu) na lineární kombinaci bázových periodických ortogonálních, ale neharmonických signálů.

## 10.3 Hadamardova transformace

Zde jsou bázovými funkcemi reálné (ne komplexní) pravoúhlé průběhy nazývané Walshovy funkce. Jejich hodnota může být pouze  $\pm 1$ . Walshovy bázové funkce jsou uspořádány podle počtu průchodů nulovou úrovni, podobně jako jsou sinusovky a kosinusovky u Fourierovy transformace seřazeny podle frekvence.

Zavádějí se Hadamardovy matice  $\mathbf{H}_{JJ}$ , kterými jsou symetrické matice o rozměru  $J \times J$ , jejichž elementy mají hodnotu  $\pm 1$ ; Pro jejich vytváření se použije rekurzivní postup. Začíná se od Hadamardovy matice rozměru dva, tj.  $\mathbf{H}_{22}$ , která je dána

$$\mathbf{H}_{22} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Hadamardova matice řádu  $2^k$  je

$$\mathbf{H}_{2^k 2^k} = \begin{bmatrix} \mathbf{H}_{JJ} & \mathbf{H}_{JJ} \\ \mathbf{H}_{JJ} & -\mathbf{H}_{JJ} \end{bmatrix}.$$

Existují i Hadamardovy matice jiných řádů než  $2^k$ , ale v digitálním zpracování obrazů se nepoužívají. Inverzní matice k Hadamardově matici má tvar

$$\mathbf{H}_{JJ}^{-1} = \frac{1}{J} \mathbf{H}_{JJ}.$$

Nyní již můžeme zavést Hadamardovu transformaci, a to dosazením Hadamardových matic za matice  $\mathbf{P}$ ,  $\mathbf{Q}$  v obecném vztahu (10.2) pro lineární integrální transformaci. Pro vstupní obraz  $\mathbf{f}$  je výsledkem Hadamardovy transformace obraz  $\mathbf{F}$  podle vztahu

$$\mathbf{F} = \mathbf{H}_{MM} \mathbf{f} \mathbf{H}_{NN}.$$

Inverzní Hadamardova transformace je

$$\mathbf{f} = \frac{1}{MN} \mathbf{H}_{MM} \mathbf{F} \mathbf{H}_{NN}.$$

Hadamardova matice obsahuje pouze dvě hodnoty, které lze snadno reprezentovat v logických elektronických obvodech. To je také příčinou, proč je Hadamardova transformace oblíbená u návrhářů hardwaru. Hadamardova transformace se také někdy nazývá Walshova-Hadamardova, protože jejími bázovými funkcemi jsou Walshovy funkce. Její použití v analýze signálů a obrazů je velmi podobné jako u Fourierovy transformace. Interpretace výsledku bude přirozenějiná.

## 10.4 Diskrétní kosinová transformace

Zavedeme diskrétní kosinovou transformaci, která se především používá při komprezi obrazu. Je základem oblíbené ztrátové komprese s názvem JPEG.

Existují čtyři definice diskrétní kosinové transformace označované DCT-I, DCT-II, DCT-III a DCT-IV [79]. Ve zpracování obrazu se nejčastěji používá DCT-II. Lze ji zapsat jako zvláštní případ obecného maticového zápisu lineární integrální transformace (10.2), a to s použitím bázových funkcí (v našem diskrétním případě vektorů), kterými jsou vzorkované kosinusovky. Předpokládejme obraz rozměru  $N \times N$  a zavedeme

$$C_{NN}(k, l) = \begin{cases} \frac{1}{\sqrt{N}} & \text{pro } l = 0, \\ \sqrt{\frac{2}{N}} \cos\left(\frac{(2k+1)l\pi}{2N}\right) & \text{pro ostatní } k, l. \end{cases}$$

Přímou, resp. inverzní kosinovou transformaci lze potom v duchu obecného maticového vztahu (10.2) zapsat jako

$$\mathbf{F} = \mathbf{C}_{NN} \mathbf{f} \mathbf{C}_{NN}^T, \quad \mathbf{f} = \mathbf{C}_{NN}^T \mathbf{F} \mathbf{C}_{NN}.$$

Zapišme pro názornost levou část předchozího výrazu pro přímou kosinovou transformaci DCT-II v ekvivalentním tvaru se sumami, ze kterého je vidět, že transformovaný obraz je lineární kombinací kosinusovek

$$F(u, v) = \frac{2c(u)c(v)}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m, n) \cos\left(\frac{2m+1}{2N}u\pi\right) \cos\left(\frac{2n+1}{2N}v\pi\right), \quad (10.21)$$

$$\text{kde } c(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{pro } k = 0, \\ 1 & \text{jinde,} \end{cases} \quad u = 0, 1, \dots, N-1, \quad v = 0, 1, \dots, N-1.$$

Poznamenejme, že diskrétní kosinovou transformaci lze rychle vypočítat s využitím rychlé Fourierovy transformace. Všech  $N$  koeficientů kosinové transformace lze vypočítat pomocí  $2N$ -bodové FFT [79].

## 10.5 Vlnková transformace (wavelets)

Vlnky (angl. wavelets) jsou bázové funkce umožňující rozložit složitější signály na lineární kombinaci jednodušších, což je podobné fourierovskému rozkladu. Vlnky mají některé další přijemné vlastnosti. Zde čtenáři poskytneme pouze základní přehled, aby se mohl na hrubé úrovni orientovat. Teorií vlnkové transformace se věnuje řada matematických monografií, např. [10].

Signál popsaný fourierovským spektrem je lokalizován z hlediska frekvence, ale není umístěn do prostoru  $x, y$ , protože se spektrum počítá z celého obrazu. Myslí se tím, že spektrum dovoluje lokalizovat frekvence, ale ne výskyt příslušných frekvencí v čase u 1D signálů nebo v rovině u 2D signálů. Například u 1D signálu může malá změna ve frekvenčním spektru (výsledku Fourierovy transformace) způsobit změny kdekoliv v časové oblasti. Pro 1D signály se tento problém obchází tím, že se Fourierova transformace počítá jen pro signál uvnitř okna.

Vlnky jsou lokalizovány v obou definičních oborech, frekvenčním i časovém. Díky tomu se lépe hodí pro analýzu signálů v různých měřítcích než jednoduché sinusovky a kosinusovky. Připomeňme, že je těžké modelovat ojedinělou špičku v signálu (vznikla například díky impulsnímu šumu) jako kombinaci signálů na nekonečně dlouhém definičním oboru. Funkce, pro něž je lokalita přirozená, se pro tento účel hodí více. Popis ostrých špiček a nespojitostí v signálu je obvykle pomocí vlnkové transformace úspornější než báze ze sinusovek a kosinusovek.

Podobně jako u Fourierovy transformace jsou i vlnky odvozeny z bázové funkce  $\Phi(x)$ , která se nazývá *matečná funkce*. Báze pro vlnkovou transformaci má tvar funkcí

$$\Phi_{(s,l)}(x) = 2^{-(s/2)}\Phi(2^{-s}x - l).$$

Součinitel  $s$  určuje šířku vlnky (z výpočetního hlediska se volí mocnina dvou) a celočíselný lokalizační index  $l$  určující pozici v časové (prostorové) oblasti. Všimněte si, že průběhy  $\Phi_{(s,l)}$  jsou "sobě podobné", jsou vybrány, aby byly ortonormální (většinou, obecně nemusí být). Tedy

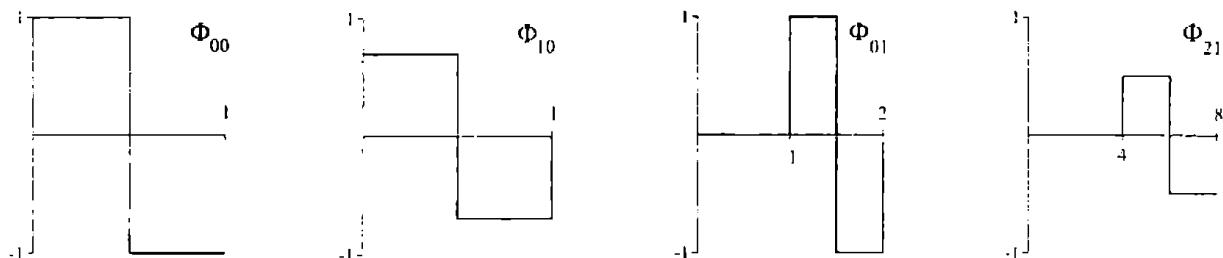
$$\int \Phi_{(s_1,l_1)} \Phi_{(s_2,l_2)} = 0 \quad \text{if } s_1 \neq s_2 \text{ or } l_1 \neq l_2.$$

Proto je možné reprezentovat jiné funkce jako lineární kombinaci  $\Phi_{(s,l)}$ .

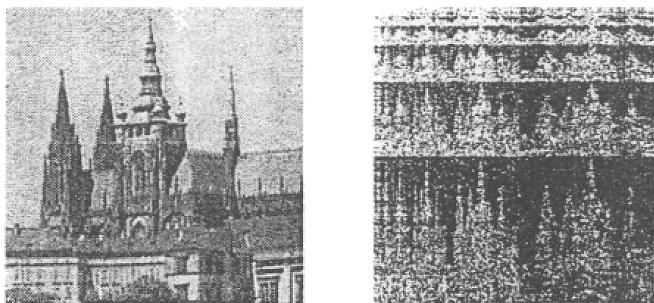
Správná volba matečné funkce je pro aplikaci vlnkové transformace klíčová. Nejjednodušší je *Haarova funkce*, kterou představuje jednoduchý skok

$$\begin{aligned}\Phi(x) &= 1 & 0 \leq x < 0.5, \\ \Phi(x) &= -1 & 0.5 \leq x < 1.\end{aligned}$$

Haarovu funkci ilustruje 10.3, ze kterého jsou patrné jevy spojené se změnou měřítka a posunem signálu. Dalšími používanými matečnými funkcemi jsou Mayerovy funkce [65] nebo funkce Ingrid Daubechiesové [10]. Příklad koeficientů vlnkové transformace pro matečné funkce podle Daubechiesové je pro ilustraci na obr. 10.4 vpravo.



Obrázek 10.3 Haarova matečná funkce  $\Phi_{00}$  a  $\Phi_{10}$ ,  $\Phi_{01}$ ,  $\Phi_{21}$ .



Obrázek 10.4 Koeficienty vlnkové transformace s matečnými funkcemi podle Daubechies. Vlevo je výchozí obraz. Vpravo jsou koeficienty. Tmavé pixely odpovídají malým hodnotám. Koeficienty v levém horním rohu odpovídají malému plošnému rozlišení. Rozlišovací schopnost se zvětšuje směrem dolů a doprava.

Vlnková transformace se ve zpracování obrazů s velkými úspěchy používá pro kompresi dat, potlačování šumu, popis obrysu objektů. Podrobnosti si lze přečíst v [65].

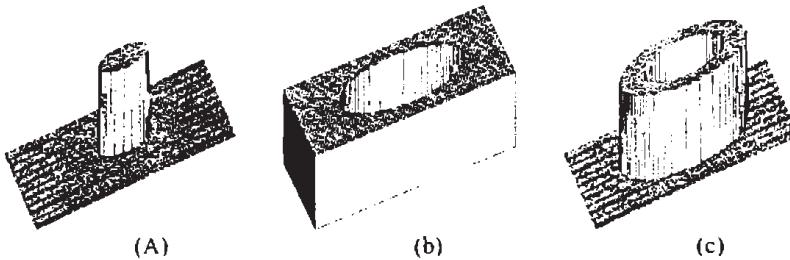
## 10.6 Další lineární integrální transformace

Existuje mnoho dalších lineárních integrálních transformací. *Paleyova* transformace a *Walshova* jsou velmi podobné popsané Hadamardově transformaci. Také používají transformační matice složené pouze z elementů  $\pm 1$  [19].

*Haarova* transformace využívá nesymetrické Haarovy matice, jejichž prvky jsou 1,  $-1$  násobené mocninami  $\sqrt{2}$  nebo 0. *Hadamardova-Haarova* transformace je kombinací obou transformací. Transformace *slant* (z anglicky skloněný, šikmý) a její modifikace nazývaná *Slant-Haarova* využívají jako bázové funkce pilovité (skloněné, šikmé) průběhy. Pro jejich výpočet existují také rychlé algoritmy. Podrobnosti o zmíněných transformacích lze najít v [77, 19].

V některých aplikacích, zejména tomografii, astronomii a holografii bývá důležité umět rekonstruovat průběh dvojrozměrné funkce (např. obrazu) z několika jejích jednorozměrných projekcí. Vznik takového obrazu lze modelovat pomocí *Radonovy* transformace [42]. *Houghova transformace*, která se často používá v segmentaci obrazu pro hledání parametricky popsaných objektů (např. přímek, kružnic), je zvláštním případem Radonovy transformace.

*Karhunenova-Loeveova* transformace používá vlastní vektory jako bázové vektory pro ortogonální rozklad kovarianční matice příslušného lineárního prostoru. Čtenář se s transformací pravděpodobně setkal v matematice nebo teorii řízení pod názvem převod matice do Jordanova kanonického tvaru. V technických oborech se také používá název metoda hlavních směrů (anglicky principal component analysis). V rozpoznávání se transformace používá pro měření informativnosti příznaků. Podrobnější popis Karhunenovy-Loeveovy transformace lze najít v [81].

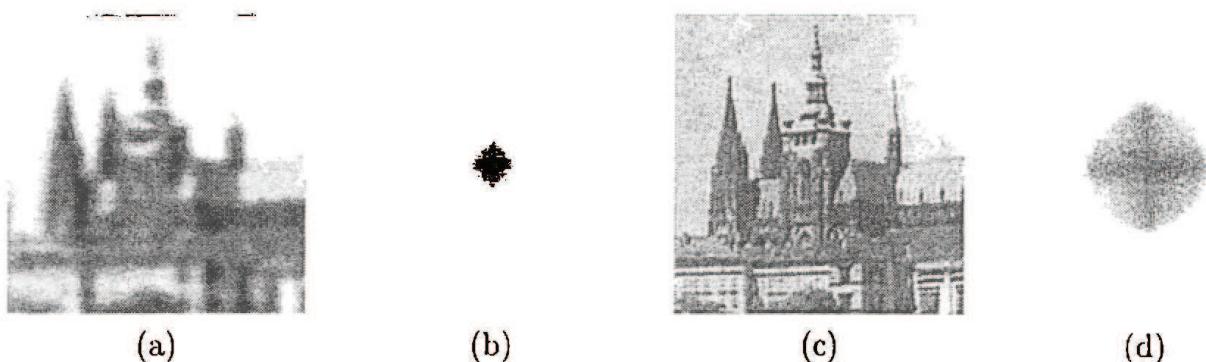


Obrázek 10.5 Frekvenční filtry zobrazené ve 3D prostoru: (a) Dolní propust, (b) horní propust, (c) pásmová propust.

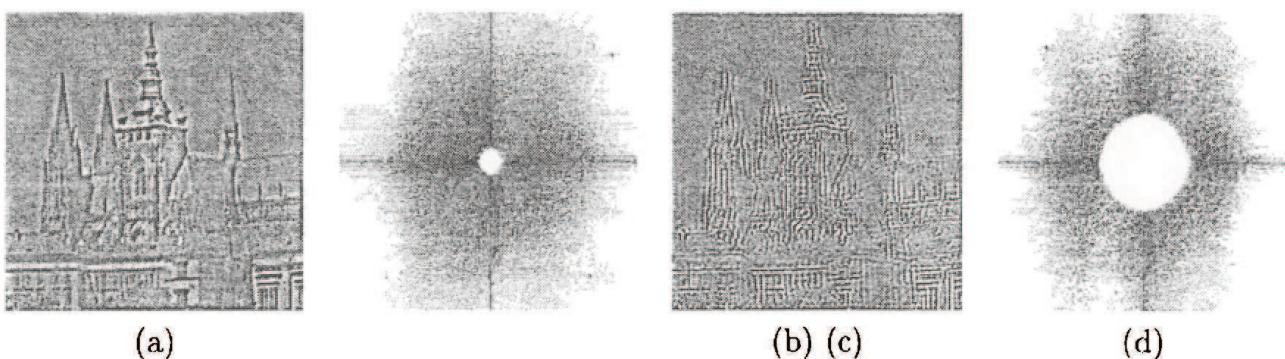
## 10.7 Použití lineárních diskrétních transformací pro zpracování obrazů

V sekci 10.2 jsme se zmínili, že Fourierova transformace umožňuje snadno vypočítat konvoluci dvou obrázků jako součin jejich Fourierových obrazů. V kapitole 11 o filtrace šumu uvidíme, jak takové konvoluční filtry použít. Fourierova transformace dovoluje interpretovat metody předzpracování obrazu založené na konvoluci jako filtraci ve frekvenční oblasti. Někdy je to názornější.

Předpokládejme opět, že  $f$  je vstupní obraz a  $F$  výsledek Fourierovy transformace. Konvoluční filtr  $h$  lze vyjádřit jako jeho Fourierův obraz  $H$ . Matice  $h$  můžeme nazývat odezvou bodového zdroje (impulsní charakteristikou filtru) a matice  $H$  frekvenční charakteristikou filtru. Zda pro filtrace použijeme  $h$  nebo  $H$  je z teoretického hlediska ihostejné. V prvním případě pracujeme v prostorové oblasti (dvojrozměrná analogie časové oblasti pro jednorozměrné průběhy) a ve druhém případě ve frekvenční. Průchod vstupního obrazu lineárním filtrem  $h$ , resp.  $H$  jsme popsali vztahy (10.18) – (10.20).



**Obrázek 10.6** Filtrace dolní propustí, tj. odstraněny vysoké frekvence; výchozí obr. 10.2: (a) Spektrum obrázku s více potlačenými vysokými frekvencemi. (b) Výsledek inverzní Fourierovy transformace spektra a. (c) Spektrum obrázku s méně potlačenými vysokými frekvencemi. (d) Výsledek inverzní Fourierovy transformace spektra c.



**Obrázek 10.7** Filtrací odstraněny nízké frekvence; výchozí obr. 10.2: (a) Spektrum obrázku s méně potlačenými nízkými frekvencemi. (b) Výsledek inverzní Fourierovy transformace spektra a. (c) Spektrum obrázku s více potlačenými vysokými frekvencemi. (d) Výsledek inverzní Fourierovy transformace spektra c.

Nyní pro názornost ukážeme jednoduché příklady dvojrozměrné filtrace ve frekvenční oblasti. Jde o dvojrozměrné příady frekvenčně selektivních filtrů z první části skripta.

*Dolní propust* odpovídá frekvenční charakteristice filtru  $H(u, v)$  s malými hodnotami v místech hodně vzdálených od počátku (malé zesílení pro vysoké frekvence) a velkými hodnotami blízko počátku, viz obr. 10.5a. Filtr přenese nízké frekvence a potlačí vysoké frekvence. Chová se stejně jako filtrace průměrováním v prostorové oblasti, tedy rozmaďává ostré hrany.

*Horní propust* odpovídá frekvenční charakteristice filtru  $H(u, v)$  s malým zesílením blízko počátku a velkým zesílením v oblasti vysokých frekvencí, tj. dále od počátku, viz obr. 10.5b.

*Pásmová propust* propustí jen frekvence v určitém pásmu, viz obr. 10.5c. Podobně jsou konstruovány filtry, které propustí jen signály o určitém směru, tj. fázi, řečeno terminologií Fourierovy transformace.

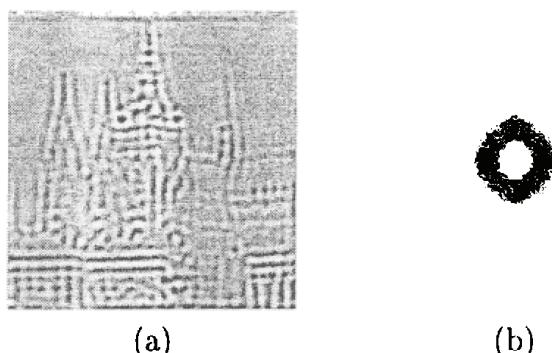
Lineární filtrace integrálními transformacemi se nejčastěji používá pro potlačení šumu, zvýraznění hran a odstranění strukturovaného šumu dobře odlišitelného od signálu ve frekvenčním spektru.

Šum obvykle má široké frekvenční pásmo, a proto ho potlačí, když jsou omezeny/odstraněny vysoké frekvence dolní propustí, viz 10.6. Naneštěstí spolu s šumem zmizí všechny jevy v obrázku odpovídající vysokým frekvencím (hrany, detaily jako tenké čáry, atd.). Obrázek je proto celkově rozmaďán.

Hrany odpovídají místům, kde se náhle mění obrazová funkce, což přispívá k vysokým

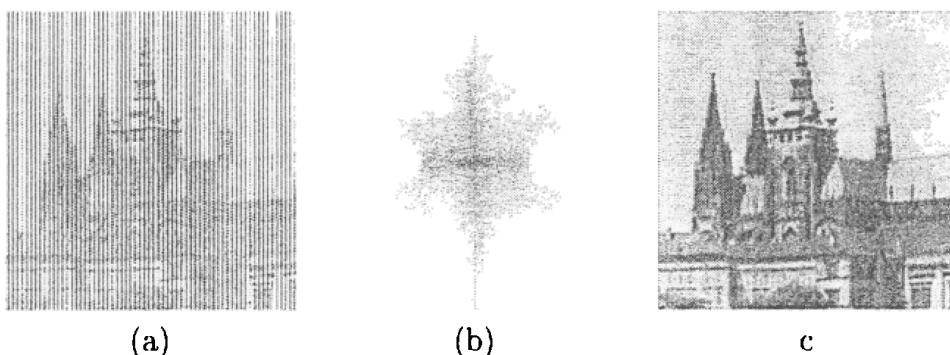
frekvencím ve spektru. Pokud vyšší frekvence relativně zesílíme, a to udělá horní propust, zvýrazní se v obrazu hrany, viz obr. 10.7.

Poměrně často se stává, že šum má výraznou strukturu ve frekvenčním spektru. Charakter šumu může být buď znám předem, nebo se může podařit odhadnout jeho vlastnosti z frekvenčního spektra zašuměného obrázku, kde se projevuje znatelnými špičkami.



**Obrázek 10.8** Filtrace pásmovou propustí; výchozí obr. 10.2: (a) Ve spektru je zachován jen pás frekvencí. Všechny nižší a vyšší frekvence jsou vyfiltrovány. (b) Výsledek inverzní Fourierovy transformace.

Ukážeme si, jak lze odstranit poruchy způsobené periodickými svislými pruhy na obrázku. Na obr. 10.9a je výchozí obrázek. Na obr. 10.9b je vidět frekvenční spektrum reprezentující filtr (pásmovou propust), který pruhy odstraní, jak je ostatně vidět z výsledku zpětné Fourierovy transformace na obr. 10.9c.

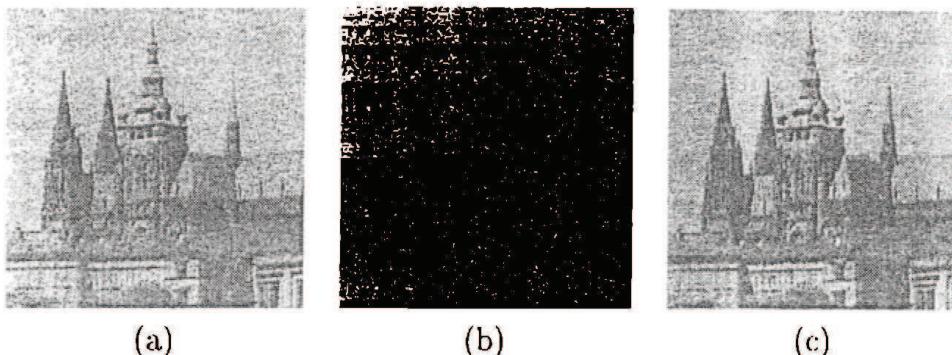


**Obrázek 10.9** Odstranění periodického šumu: (a) Výchozí obraz. (b) Frekvenční spektrum použité pro rekonstrukci. Filtr odstraní oblasti ve spektru, ve kterých leží odezva odpovídající periodickým pruhům. (c) Výsledek filtrace.

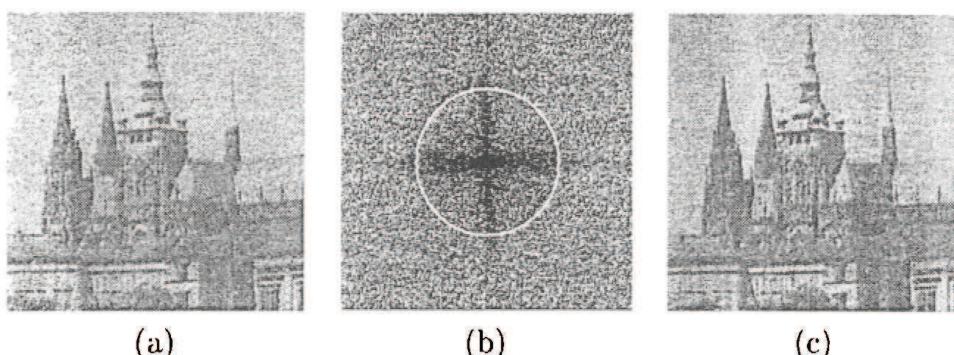
V sekci 10.5 jsme uvedli vlnkovou (waveletovou) transformaci. Její hlavní výhodou je lokalizované chování umožňující změnit/odstranit jednotlivé koeficienty transformace, aniž by to ovlivnilo celý rekonstruovaný obraz. Filtrace pomocí vlnkové transformace umožňuje potlačit náhodný šum, aniž by současně došlo k rozmažání, které je charakteristické pro ostatní lineární integrální transformace.

Ukažme si chování vlnkové transformace při filtrace šumu na příkladě. Obrázek Hradčan, viz obr. 10.4a, byl zatížen náhodným gaussovským šumem s nulovou střední hodnotou, viz obr. 10.10a. Ze zašuměného obrazu jsou spočítány koeficienty vlnkové transformace. Z nich je ponecháno jen 15% s největší hodnotou. Ostatní jsou potlačeny. Příslušné koeficienty jsou ukázány na obr. 10.10b. Výsledek inverzní vlnkové transformace je vidět na obr. 10.10c.

Pro porovnání byl na stejný výchozí zašuměný obrázek, viz obr. 10.10a nebo obr. 10.11a použita filtrace ve Fourierově spektru, které je celé zobrazeno na obr. 10.11b. Při filtrace jsou

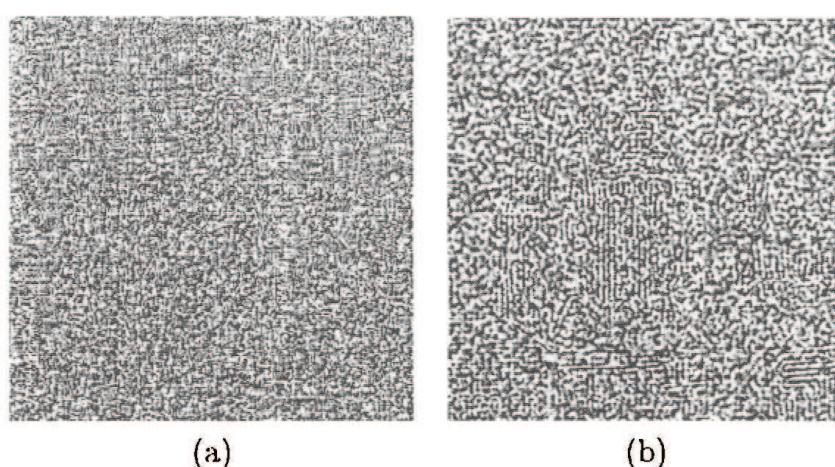


**Obrázek 10.10** Filtrace šumu vlnkovou transformací: (a) Výchozí zašuměný obrázek. (b) Koeficienty vlnkové transformace po filtraci (bylo odstraněno 85% koeficientů transformace). Odstraněné koeficienty jsou zobrazeny černě. Porovnejte s koeficienty vlnkové transformace původního nezašuměného obrázku na obr. 10.4. (c) Výsledek inverzní vlnkové transformace.



**Obrázek 10.11** Filtrace šumu potlačením vysokých frekvencí ve Fourierově spektru: (a) Výchozí zašuměný obrázek. (b) Fourierovo spektrum obrázku a. Spektrum vně bílé kružnice je potlačeno. (c) Výsledek inverzní Fourierovy transformace.

potlačeny frekvence vně kružnice. Výsledný filtrovaný obraz vznikl zpětnou Fourierovou transformací, viz obr. 10.11c. Porovnáním obr. 10.10c s obr. 10.11c je vidět, že vlnková transformace filtruje lépe – malé detaily nejsou rozmažány.



**Obrázek 10.12** Rozdílové obrázky pro porovnání filtrace šumu vlnkovou a Fourierovou transformací. (a) Absolutní diference mezi původním (zašuměným) obrázkem a výsledkem vlnkové filtrace, obr. 10.10c. (b) Absolutní diference mezi původním (zašuměným) obrázkem a výsledkem filtrace ve Fourierově spektru, viz obr. 10.11c.

Aby bylo možné tento subjektivní pocit objektivizovat, jsou na obr. 10.12 ukázány rozdílové snímky. Všimněte si poruch blízko ostrých jasových rozhraní po Fourierově filtrace. Vlnková filtrace těmito neduhy netrpí.

Na závěr kapitoly o dvojrozměrných lineárních diskrétních transformacích poznamenejme, že filtrace obrazu není jedinou oblastí jejich použití. Další aplikací je komprese obrazu, popis objektů v obrazu, např. textury, a teorie rozpoznávání, kde jsou používány pro výběr a uspořádání příznaků.

# 11. Filtrace šumu, poruch a detekce hran

## 11.1 Statistický princip filtrace šumu

Nechť je každý obrazový bod zatížen náhodným aditivním šumem  $\nu$ , který je na obrazové funkci nezávislý, má nulovou střední hodnotu a směrodatnou odchylku  $\sigma$ . Více realizací takových obrazů můžeme získat  $n$ -násobným sejmutím statické scény za stejných podmínek. Z každého sejmutého obrazu vybereme bod o stejných souřadnicích, označíme ho  $g_i$ ,  $i = 1 \dots n$ , kde index  $i$  ukazuje, ke kterému obrazu bod patří. Odhad správné hodnoty můžeme získat jako aritmetický průměr bodů  $g_1, \dots, g_n$  se šumem  $\nu_1, \dots, \nu_n$ ,

$$\frac{g_1 + \dots + g_n}{n} + \frac{\nu_1 + \dots + \nu_n}{n}.$$

Pravý sčítanec předchozího výrazu popisující vliv šumu po průměrování ukazuje, že jde opět o náhodnou veličinu s nulovou střední hodnotou a menší směrodatnou odchylkou o velikosti  $\sigma/\sqrt{n}$ . Je vidět, že se směrodatná odchylka zmenšila  $\sqrt{n}$ -krát. Obdobný jednorozměrný případ byl popsán v kapitole 5.

Předchozí úvaha je obvyklá ve statistice, kde je cílem odhadnout průměr nějaké populace. Z populace pořídíme náhodný výběr a spočítáme jemu odpovídající (výběrový) průměr. Pokud bychom náhodný výběr mnohokrát opakovali a počítali příslušné výběrové průměry, dostali bychom rozdělení výběrových průměrů. Shrňme neformálně tyto vlastnosti:

- Průměr rozdělení výběrových průměrů se rovná průměru populace.
- Rozdělení výběrových průměrů má menší rozptyl než původní populace. Je-li standardní odchylka populace rovna  $\sigma$ , potom je standardní odchylka výběrových průměrů rovna  $\sigma/\sqrt{n}$ .
- Nepřekvapí nás, že je-li rozdělení v populaci normální, je normální i rozdělení výběrových průměrů. Důležitější je důsledek *centrální limitní věty*. Tvrdí, že jsou-li k dispozici dostatečně velké náhodné výběry, potom je rozdělení výběrových průměrů blízké k normálnímu rozdělení, i když populace normální rozdělení nemá. Proto hraje normální rozdělení ve statistice tak významnou roli.
- Pro aplikace je podstatné, že není potřebné generovat velké innožství náhodných výběrů. Stačí pořídit jedený dosti velký náhodný výběr a díky centrální limitní větě víme, jaké je rozdělení výběrových průměrů, aniž je musíme generovat. Co lze považovat za dostatečně velký výběr? Ve statistice bývá považováno za dolní hranici 30 pozorování. Míra nejistoty hodnoty populačního parametru (zde jsme mluvili jen o průměru) se vyjadřuje intervalém spolehlivosti. Zájemce odkazujeme na statistickou literaturu.

## 11.2 Lokální filtrace

Předmětem našeho zájmu v této sekci budou metody filtrace (předzpracování) obrazu, které k výpočtu nové hodnoty pixelu využívají malé okolí  $\mathcal{O}$  reprezentativního pixelu (ve smyslu

právě zpracovávaného). Přístup je založen na představě, že se celý obraz systematicky (např. po řádcích) prochází. Kolem reprezentativního bodu je zkoumáno malé okolí  $\mathcal{O}$ , často malý obdélník. Výsledek analýzy je zapsán do výstupního obrazu jako hodnota reprezentativního pixelu.

Podle účelu se metody lokálního předzpracování rozdělují do dvou skupin. První skupina, *vyhlazování*, usiluje o potlačení šumu a osamocených fluktuací hodnot obrazové funkce. Myšlenkově jsou metody příbuzné dolnafrekvenčním propustem z minulé kapitoly. Druhá skupina, *detekce hran*, též gradientní operátory, se snaží z hodnot v okolí reprezentativního pixelu odhadnout derivaci obrazové funkce. Metody detekce hran jsou příbuzné s hornofrekvenčními propustmi probranými v minulé kapitole. Je vidět, že vyhlazování a detekce hran jsou (alespoň ve své lineární podobě) v protikladu. Proto jsou navrhovány nelineární metody, které např. vyhlazují a přitom jsou šetrné k hranám a detailům v obrazu.

Jiným hlediskem pro třídění metod předzpracování jsou matematické vlastnosti příslušné transformace. Lze rozlišit *lineární* a *nelineární* metody. Připomeňme, že lineární operace počítají hodnotu ve výstupním obrazu  $g(x, y)$  jako lineární kombinaci hodnot vstupního obrazu  $f$  v malém okolí reprezentativního pixelu  $(x, y)$ .

Příspěvek jednotlivých pixelů v okolí  $\mathcal{O}$  je vážen koeficienty  $h$  podle

$$f(x, y) = \sum_{(m,n) \in \mathcal{O}} h(x - m, y - n) g(m, n). \quad (11.1)$$

Rovnice (11.1) popisuje diskrétní *konvoluci* s jádrem  $h$ , kterému se také říká *konvoluční maska*. Často se používá obdélníkové okolí  $\mathcal{O}$  s lichým počtem řádků a sloupců, a tak může reprezentativní bod ležet uprostřed konvoluční masky.

Předzpracování obvykle nepoužívá specifické znalosti o obrazu. Je také těžké takovou znalost při zpracování odvodit, protože známé okolí  $\mathcal{O}$  právě zpracovávaného bodu je malé. Je-li znalost o obrazu k dispozici, potom toho algoritmus zpracování může s výhodou využít. Někdy takovou znalostí bývají známé statistické parametry šumu.

Je důležité si uvědomit, že v průběhu předzpracování (stejně jako při každém jiném zpracování) nezískáme z hlediska Shannonovy teorie žádnou novou informaci. Tu bychom měřili entropií. Můžeme jen některou informaci potlačit nebo zvýraznit. Potřebuje-li zvýšit celkový informační obsah obrazu, potom musíme zlepšit postup jeho pořízení.

Konvoluční lokální operace (filtry) jsou použitelné pro vyhlazování, detekci hran nebo čar. Pro zrychlení výpočtu těchto operací lze poměrně jednoduše vytvořit specializované technické prostředky. Zjednodušení hardwaru (také zrychlení výpočtů) je zvláště podstatné pro speciální tvary filtrů, jako jsou separabilní a rekurzivní filtry, o nichž budeme mluvit v části 11.3.1.

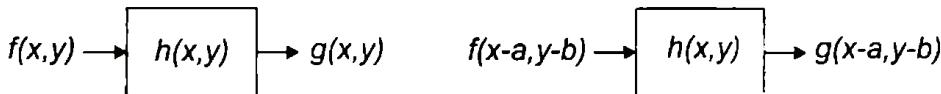
### 11.3 Lokální vyhlazování obrazu

Nejsnadnější je vyhlazování náhodného šumu, když máme k dispozici několik obrazů téže předlohy, které se liší právě šumem. Potom je účinné průměrovat hodnotu pixelu o stejných souřadnicích přes více obrázků. Tento přístup jsme již rozebrali v sekci 11.1. Poznamenejme jen, že zde problémy s rozmařáváním přirozeně nevznikají.

Máme-li k dispozici jen jediný obraz téže předlohy, potom nezbývá než se spolehnout na obvykle značnou nadbytečnost údajů v obrazu. Sousední pixely mají převážně tutéž nebo blízkou hodnotu jasu. Hodnotu obrazových elementů zkreslených šumem potom můžeme opravit na základě analýzy hodnot jasu v jeho vybraném okolí. Hodnota reprezentativního pixelu (tj. jeho jas) je nahrazena hodnotou typického reprezentanta mezi hodnotami v okolí nebo kombinací několika hodnot.

### 11.3.1 Lineární metody vyhlazování

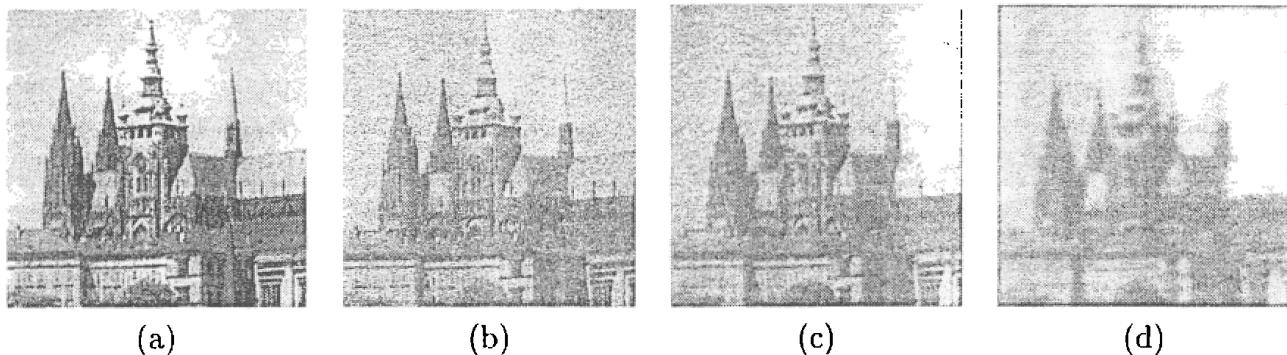
Lineární metody vyhlazování vypočítávají novou hodnotu reprezentativního pixelu jako lineární kombinaci hodnot ve zkoumaném okolí. Pro digitální snímky lze v prostorové oblasti lineární kombinaci vyjádřit jako diskrétní konvoluci (11.1). Jednotlivé lineární filtry se liší váhami v lineární kombinaci, které jsou dány příslušnou konvoluční maskou  $h$  (dvojrozměrnou impulsní odevzou).



Obrázek 11.1 Ilustrace lineárního filtru (vlevo) a lineárního prostorově invariantního filtru (vpravo).

Ve zpracování obrazu se používá zvláštní třída lineárních filtrov. Říká se jim *prostorově invariantní filtrov* (někdy se používá i název homogenní filtrov), protože se chování filtru nemění při změně polohy v obrázku, obr. 11.1. V teorii jednorozměrných signálů je analogickou vlastností časová invariantnost filtrov. Filtr opírající se o představu postupné konvoluce s malou maskou podle vztahu (11.1) je prostorově invariantní.

Při zpracování skutečných obrazů je předpoklad linearity narušen díky tomu, že hodnota obrazové funkce (jas, intenzita) je nezáporná a omezená. Podobně jsou obrazy ohrazeny v prostoru, a tak předpoklad prostorové invariantnosti platí jen pro omezené posuny konvolučních masek. Jelikož není k dispozici jednoduchá teorie, která by popisovala zmíněné potíže, používá se obvykle lineární přístup a dodatečně se zajišťuje, aby na okrajích obrazu nebo v mezích rozsahu hodnoty obrazové funkce nevznikaly problémy.



Obrázek 11.2 Vyhazování šumu s Gausovým rozdělením: (a) výchozí obraz  $256^2$ , (b) uměle přidán Gausovský šum v jasu, (c) výsledek obyčejného průměrování v okně  $3 \times 3$ , (d) výsledek obyčejného průměrování v okně  $7 \times 7$ .

Základní metodou vyhlazování obrazu je *obyčejné průměrování*, kde každému bodu přiřadíme nový jas, který je aritmetickým průměrem původních jasů ve zvoleném okolí. Odpovídající konvoluční maska  $h$  pro okolí  $3 \times 3$  je

$$h = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (11.2)$$

Někdy se zvětšuje váha středového bodu masky nebo jeho 4-sousedů. Následující vztahy ukazují

tyto masky pro okolí  $3 \times 3$ . Větší masky se vytvářejí analogicky:

$$h = \frac{1}{10} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad h = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (11.3)$$

Ukažme si výsledky filtrace obyčejným průměrováním na skutečných obrázcích, viz obr. 11.2. Schwálme jsme zvolili malé prostorové rozlišení obrázků  $256^2$ , aby vynikl diskrétní charakter operací. Základní nevýhodou praktického použití obyčejného průměrování je rozmařávání hran v obraze.

S touto nepříjemností jsme se setkali již při lineární filtrace ve frekvenční oblasti. Budeme-li uvažovat Fourierovu transformaci, potom je konvoluce ekvivalentně vyjádřena jako součin Fourierova obrazu signálu (zde obrázku) a Fourierova obrazu konvolučního filtru, viz rovnice (10.19). Lineárnímu vyhlazování v prostorové oblasti potom odpovídá frekvenční filtrace dolnopropustným filtrem.

S ohledem na nežádoucí rozmařávání se obyčejné průměrování většinou používá jako pomocná metoda pro výpočet střední hodnoty jasu. Tento mezivýsledek je potom použit v důmyslnějších nelineárních vyhlazovacích metodách.

Z hlediska úspory výpočtů existuje důležitý zvláštní případ konvolučních filtrov, tzv. *separabilní filtrov*. Jde o případ, kdy lze konvoluční masku v  $p$ -rozměrném okolí, obvykle  $p = 2:3$  rozložit na součin jednorozměrných masek. Jinak řečeno, lze rozložit vícerozměrný systém bázových ortogonálních funkcí lineární integrální transformace na součin jednorozměrných ortogonálních funkcí. Princip je zřejmý z následujícího příkladu.

**Příklad 11.1 Separabilní filtr.** Binomický 2D filtr rozměru  $5 \times 5$  (je složen z binomických čísel, která vznikají jako součet dvou čísel v předchozím řádku Pascalova trojúhelníku) lze dekomponovat následovně jako součin dvou jednorozměrných vektorů  $h_1, h_2$

$$\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \begin{bmatrix} & h_2 & \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}.$$

Nechť rozměr konvoluční masky je  $2N+1$ . Konvoluci podle vztahu (11.1) můžeme s uvažováním speciálních vlastností separabilního filtru přepsat

$$g(x, y) = \sum_{k=-N}^N \sum_{l=-N}^N h(k, l) f(x + i, y + l) = \sum_{k=-N}^N h_1(k) \sum_{l=-N}^N h_2(l) f(x + i, y + l).$$

Pro přímý výpočet konvoluce bez dodatečních výpočetních triků by bylo třeba  $25$  násobení a  $24$  sčítání pro každý pixel. Při použití separabilního filtru stačí  $10$  součinů a  $8$  součtů.

Ještě výraznější by byla úspora v případě konvoluční filtrace 3D obrázku (např. z tomografu). Pro konvoluční jádro rozměru  $9 \times 9 \times 9$  by na každý voxel bylo potřeba  $729$  součinů a  $728$  součtů, zatímco pro separovatelný filtr stačí  $27$  součinů a  $24$  součtů na voxel. ▲

Významnou skupinou prostorově invariantních filtrov jsou *rekurzivní filtrov*. Používají se hojně zejména pro 1D signály, jak jsme viděli v sekci 4.1. Při běžné konvoluci (11.1) se vypočtené hodnoty při posunu konvoluční masky ukládají do výstupního obrazu a pro vstup se používají původní nezměněné hodnoty. Pokud jsou jako vstupní hodnoty pro konvoluci použity hodnoty

vypočtené v předchozí poloze masky (jen její části, která je již nafiltrována novými hodnotami), jde právě o rekurzivní filtr. Pro názornost si situaci připomněme v 1D případě. Nechť  $\dots, g_n, \dots$  je vstupní posloupnost a  $h_0, \dots, h_r$  je maska filtru. Rekurzivní filtr je dán

$$g'_n = h_0 g_n + \sum_{i=0}^r h_i g'_{n-i}.$$

Jak je vidět z předchozího vztahu, v konvoluci se kombinují dosud nefiltrované a také již filtrované hodnoty, na což lze nahlížet jako na zavedení zpětné vazby. Rekurzivní filtry z principu fungují v jednom daném směru. V 1D filtroch je přirozený směr určen rostoucím časem. Zvláštní třída *kauzálních filtrov* používá pro stanovení výstupní hodnoty pouze současnou a minulé hodnoty signálu. Pro uchování minulých hodnot signálů je potřebná paměť, a proto jde o dynamický systém. Čtenář již studoval dynamické systémy v jiných předmětech (např. teorii obvodů, teorii řízení) a také v první části tohoto skripta.

Pro 1D linerární rekurzivní filtry platí vše, co bylo řečeno o filtroch s konečnou (FIR) a nekonečnou (IIR) impulsní odezvou v kapitole 4. Shrňme některé jejich důležité vlastnosti. Rekurzivní filtry mohou být nestabilní, a to na rozdíl od nerekurzivních konvolučních filtrov, které jsou stabilní vždy. Při návrhu rekurzivních filtrov je požadavek na stabilitu jedním z požadavků. Rekurzivní filtry jsou výpočetně efektivnější než nerekurzivní. Pro požadovanou odezvu stačí filtr popsat méně parametry než v nerekurzivním případě (díky zpětné vazbě). Kauzální rekurzivní filtry se vyznačují záporným fázovým posunem (zpožděním), které nelineárně závisí na pořadovém čísle ortogonální bázové funkce (u Fourierovy transformace bychom říkali harmonické složce).

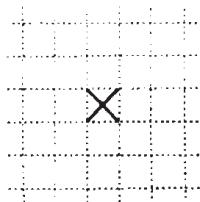
Použití rekurzivních filtrov pro 2D obrazy je spojeno s několika problémy. Pro obrazy neexistuje přirozený směr, jakým byl čas pro 1D signály. Není principiální důvod, proč by některý směr měl být preferován. Ve zpracování obrazů se dává přednost nekauzálním filtrov s nulovým fázovým posunem, a to proto, aby se v obraze neposouvaly hrany nebo jiné body významné pro další analýzu. Pro 1D rekurzivní filtry je vypracovaná jasná teorie. Pro vícerozměrné filtry tomu tak není, protože zobecnění Z-transformace do více dimenzí není snadné. Působí to základní problémy při návrhu rekurzivních filtrov a studiu jejich stability. Osvědčenou cestou návrhu je kaskádní řazení jednoduchých filtrov [41].

### 11.3.2 Nelineární metody vyhlazování

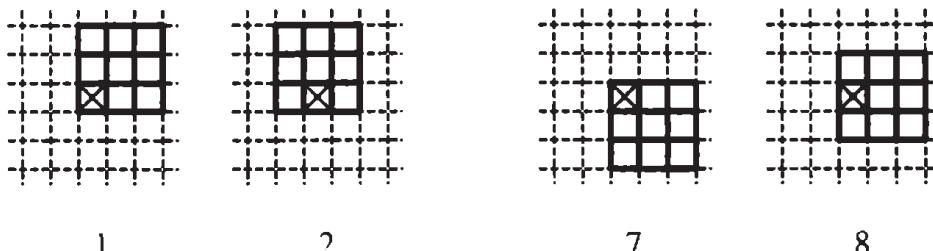
Potíže s rozmařáváním hraničástečně eliminují nelineární filtrační metody. Jejich společným principem je to, že se v analyzovaném okolí  $\mathcal{O}$  snaží najít jen tu jeho část (oblast o zhruba konstantnímu jasu), do které reprezentativní bod patří. Jen pixely této oblasti se využijí pro hledání jasové hodnoty (např. aritmetickým průměrem nebo výběrem hodnoty jasu jednoho pixelu), která bude reprezentovat celé okolí  $\mathcal{O}$  ve výstupním obrázku. Představme si, že např. v obdélníkovém okolí  $\mathcal{O}$  rozměru  $5 \times 5$  leží roh tmavého objektu, viz obr. 11.3. Je přirozené hledat reprezentanta jen uvnitř objektu, což předpokládá nelineární operaci - výběr.

Jako příklad třídy metod vyhlazování, které filtruji jen ve vybraném okolí, uvedeme metodu rotující masky. Kolem reprezentativního bodu rotuje malá maska, v našem příkladě na obr. 11.3 je maskou čtverec  $3 \times 3$ . Možných poloh masky je 8. V každé masce se spočte rozptyl jasu. Za homogenní okolí reprezentativního bodu se vybere maska s nejmenším rozptylem. Nová hodnota reprezentativního bodu může být dána aritmetickým průměrem hodnot ve vybrané masce. Metoda nerozmazává hrany obrazu, má dokonce mírně ostřící charakter.

Metodu rotující masky lze použít iterativně. Tento proces poměrně rychle konverguje do



**Obrázek 11.3** Na jasových rozhraních je potřeba filtrovat jen v pixlech příslušejících objektu. Křížek označuje reprezentativní pixel.



**Obrázek 11.4** Osm možných poloh rotující masky. Je možné volit i jiné tvary masky než čtverce.

stabilního stavu, kdy se již obraz dále nemění. Velikost a tvar masky ovlivňuje rychlosť konvergencie. Čím menší jsou masky, tím menší jsou změny a tím více iterací je potřeba.

Jako příklad další používané nelineární metody uvedeme *obyčejné průměrování s omezením změn jasu*, které využívá obyčejného průměrování, ale brání rozmažání hran povolením pouze určitého intervalu diference mezi jasem původního bodu a výsledkem průměrování. Když je diference menší než předem zvolený práh, použije se výsledek průměrování. V opačném případě se původní hodnota nezmění. Metoda se používá pro opravu velkoplošných chyb bez vlivu na zbytek obrazu nebo pro jednoduché vyhlazení bez poškození hran.

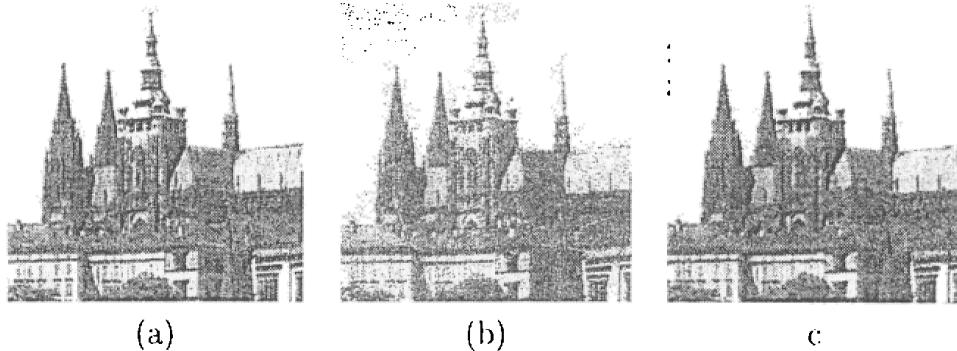
Nyní se budeme věnovat nelineárním filtrům, jejichž původ je v robustní statistice. Upozorněme na hlavní potíž lineárního odhadu správné hodnoty z jejich zašuměných realizací. Pro názornost si představme, že odhadujeme správný jas z instancí v okně rozměru  $3 \times 3$ . Nechť se osm hodnot jen málo liší od správné hodnoty a jedna hodnota je (např. díky hrubé chybě při měření) vychýlená, řekněme blížící se k  $\infty$ . Aritmetický průměr, tedy lineární odhad, vyjde také  $\infty$  a bude vychýlený. Tento problém je vlastní všem lineárním odhadům, například i často používané approximaci pomocí metody nejmenších čtverců. Představme si její jednoduchou variantu – prokládání množiny bodů přímkou. Pokud jeden bod bude značně vychýlený, bude vychýlená i approximující přímlinka.

Základem metod robustní statistiky je snaha najít v náhodné populaci vychýlené hodnoty, vyloučit je z uvažování a zbylé hodnoty nahradit nějakou typickou hodnotou. Uvedeme příklad hodnocení sportovního výkonu, který neformálně používá podobnou myšlenku.

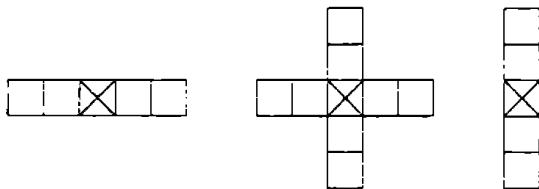
**Příklad 11.2 Robustní statistika, hodnocení skoku na lyžích.** Skoky na lyžích se hodnotí podle bodů, které oceňují jak délku skoku tak i skokanův styl. Přispěvek k bodovému hodnocení za délku skoku je objektivní a hodnocení stylu je značně subjektivní. Skoky hodnotí několik rozhodčích. Aby se vyloučila možnost rozhodčích podpořit "svého" závodníka nebo naopak "cizího" závodníka poškodit, nebore se v úvahu hodnocení rozhodčími s nejvyšším a nejnižším počtem bodů (tj. vyloučení možných vychýlených hodnot). Z hodnocení zbylých rozhodčích se vypočte aritmetický průměr, což odpovídá nalezení typické hodnoty. ▲

Ve statistice ztělesňují myšlenku robustních statistik *výběrové kvantily*. Jejich nejjednodušším případem a ve zpracování obrazů nejčastěji používaným případem je *medián M*, který již byl

zmíněn v kapitole 5.1.2. Nechť je  $x$  náhodnou veličinou. Medián  $M$  je hodnota, pro kterou je pravděpodobnost jevu  $x < M$  rovna jedné polovině. Výpočet mediánu je pro diskrétní obrazovou funkci jednoduchý. Stačí uspořádat vzestupně hodnoty jasu v lokálním okolí a medián určit jako prvek, který je uprostřed této posloupnosti. Aby se snadno určil prostřední prvek, používají se posloupnosti s lichým počtem prvků. Pokud se v obrázcích používá lokální čtvercové okolí, jsou jeho rozměry liché, např.  $3 \times 3$ ,  $5 \times 5$ , atd. Výpočet ještě urychlí skutečnost, že k nalezení mediánu stačí částečné uspořádání posloupnosti.



**Obrázek 11.5** Filtrace mediánem: (a) Výchozí obraz. (b) Umělé porušení impulsním šumem. (c) Výsledek filtrace mediánem v okně  $3 \times 3$ .



**Obrázek 11.6** Okolí používané pro zachování svislých a vodorovných čar, např. při mediánové filtrace.

Metoda filtrace mediánem stanoví jas výsledného bodu jako medián určený z hodnot jasu bodů v lokálním okolí (např.  $3 \times 3$ ) vstupního obrazu. Metoda redukuje stupeň rozmaření hran a dobré potlačuje impulsní šum. Ukažme si výsledky filtrace mediánem na příkladě. Původní obrázek (obr. 11.5a) byl uměle poškozen impulzním šumem (obr. 11.5b). Výsledek filtrace mediánem s oknem rozměrem  $3 \times 3$  je na obr. 11.5c. Filraci mediánem lze použít iterativně. Hlavní nevýhodou filtrace mediánem v obdélníkovém okolí je to, že porušuje tenké čáry a ostré rohy v obraze. Tento nedostatek se dá obejít, když se použije jiný tvar okolí, např. podle obr. 11.6.

## 11.4 Hledání hran

### 11.4.1 Pojem hrany a princip gradientních operátorů

Výsledky neurofyzioligického a psychologického výzkumu ukazují, že pro vnímání člověka jsou velmi důležitá místa v obraze, kde se náhle mění hodnota jasu. Odpovídající pixely se nazývají hrany. Například u perokresby, kdy je potřebné silně generalizovat a zaznamenat scénu velmi zjednodušeně, budou linie odpovídat právě náhlým jasovým změnám. K automatickému nalezáni takových významných míst v obraze slouží právě lokální předzpracování – hledání hran.

Hrana v obraze je dáná vlastností obrazového elementu a jeho okolí. Je určena tím, jak náhle se mění hodnota obrazové funkce  $f(x, y)$ . Matematickým nástrojem pro studium změn

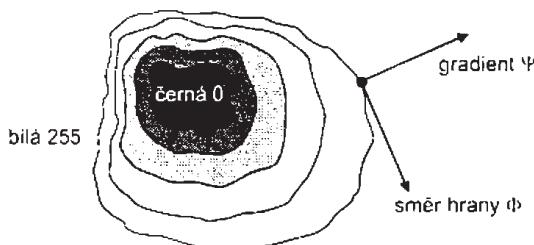
funkce dvou proměnných jsou parciální derivace. Změnu funkce udává její *gradient*, vektorová veličina  $\nabla$ , určující směr největšího růstu funkce (směr gradientu) a strniště tohoto růstu (velikost, modul gradientu). Pixely s velkým modulem gradientu se nazývají hranami.

Pro spojitu obrazovou funkci  $f(x, y)$  jsou velikost gradientu  $|\nabla f(x, y)|$  a směr gradientu  $\psi$  dány vztahy

$$|\nabla f(x, y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}, \quad (11.4)$$

$$\psi = \arg\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right), \quad (11.5)$$

kde  $\arg(x, y)$  je úhel (v radiánech) mezi souřadnicí osou  $x$  a radiusvektorem k bodu  $(x, y)$ .

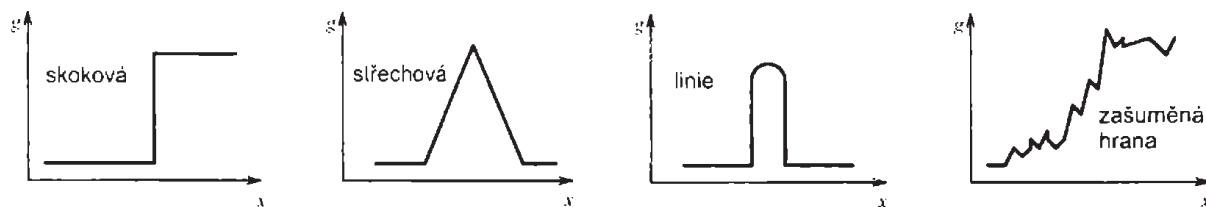


Obrázek 11.7 Směr hrany  $\Phi$  je kolmý na směr gradientu  $\Psi$ .

Někdy nás zajímají místa v obraze s významnějšími hranami (místa, kde se jas prudce mění) bez ohledu na směrové vlastnosti hran. Často se používá všeobecný lineární Laplaceův operátor - *Laplacián*  $\nabla^2$ , který vychází z druhých parciálních derivací. Pro monotónně rostoucí jasovou funkci  $g$  v příslušném okolí je Laplacián  $\nabla^2$  je nulový, kde je velikost gradientu  $|\nabla|$  maximální.

$$\nabla^2 g(x, y) = \frac{\partial^2 g(x, y)}{\partial x^2} + \frac{\partial^2 g(x, y)}{\partial y^2}. \quad (11.6)$$

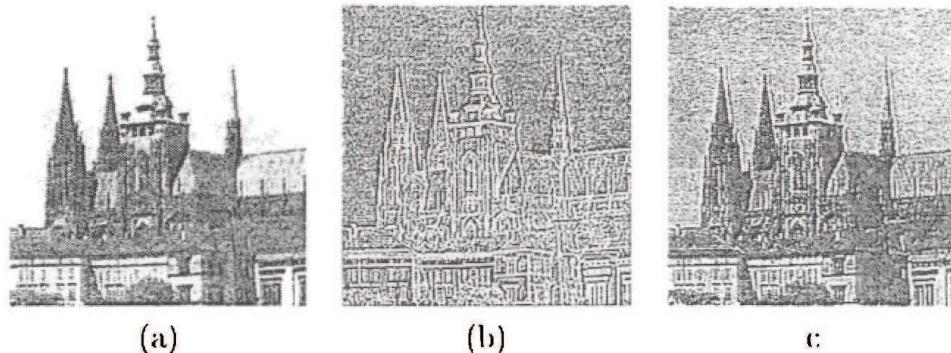
Hrany nalezené v obraze lokálními operátory se někdy používají pro hledání hranic objektů. Za předpokladu, že objektu odpovídá oblast homogenního jasu, jsou body hranice právě pixely s vysokou hodnotou gradientu. Hranové pixely se spojují do hranic, a proto se směr hrany  $\Phi$  někdy definuje jako kolmý na směr gradientu  $\Psi$ . Myšlenku ilustruje obr. 11.7. Hranové pixely se při takové volbě orientace hrany mohou přirozeně spojovat do hranic.



Obrázek 11.8 Jasové profily nejběžnějších hran.

Hrany lze dobře třídit podle jednorozměrného jasového profilu ve směru gradientu v daném pixelu. Typické příklady jsou na obr. 11.8. První tři profily, tj. skoková hrana, střechová hrana, tenká linie, jsou idealizované. Poslední profil odpovídá zašuměné hraně, kterou lze najít v reálném obrázku.

Gradientní operátory lze využít i pro *ostření obrazu*. Cílem ostření obrazu je upravit obraz tak, aby v něm byly strmější hranы. Zostřený obraz bude pozorovat člověk. Z hlediska plošných



**Obrázek 11.9** Ostření obrazu: (a) Výchozí obraz. (b) Odezva Laplaceánu v 8-okolí. (c) Výsledek ostření podle (11.7), kde  $S(x, y)$  je Laplaceán s  $C = 0,4$ .

frekvencí odpovídá ostření zdůraznění vysokých frekvencí. Pro obraz  $f$ , který je výsledkem ostření obrazu  $g$ , lze napsat

$$f(x, y) = g(x, y) - C S(x, y), \quad (11.7)$$

kde  $C$  je kladný součinitel udávající požadovanou sílu ostření a  $S(i, j)$  je operátor udávající strmost zmiňny obrazové funkce v příslušném bodě. Strmost může být dánou modulem gradientu nebo Laplaceánu. Ostření obrazu s využitím Laplaceánu ilustruje obr. 11.9.

Ostření obrazu lze také interpretovat ve frekvenčním spektru jako zdůraznění vysokých frekvencí. Víme již, že výsledek Fourierovy transformace je lineární kombinací harmonických průběhů, např.  $\sin(nx)$ ,  $n = 1, 2, \dots$ . Derivace harmonické bude  $n \cos(nx)$ . Je vidět, že čím vyšší frekvence (zde vyšší  $n$ ), tím větší je amplituda její derivace. Jde o další zdůvodnění, proč gradientní operátory zdůrazňují hrany.

Ostření obrazu se používá, když se má dosáhnout kontrastnějšího obrazu, a to buď na displeji, nebo při tisku. Na běžné laserové tiskárně a v řadě polygrafických technik (např. při sítotisku) se odstíny jasu vytvářejí polotónováním, tj. větší či menší hustotou černých. Před tiskem pomocí polotónování bývá ostření obrazu nezbytné. Příslušný modul je součástí každého programového balíku pro publikování pomocí počítače (též DTP z angl. Desk Top Publishing).

V digitalizovaném obrazu approximujeme parciální derivace diferencemi

$$\begin{aligned} \Delta_x g(x, y) &= g(x, y) - g(x - n, y), \\ \Delta_y g(x, y) &= g(x, y) - g(x, y - n), \end{aligned} \quad (11.8)$$

kde  $n$  je malé celé číslo, obvykle 1. Diferenci lze vyjádřit i symetricky

$$\begin{aligned} \Delta_x g(x, y) &= g(x + n, y) - g(x - n, y), \\ \Delta_y g(x, y) &= g(x, y + n) - g(x, y - n), \end{aligned} \quad (11.9)$$

ale obvykle se nepoužívá, protože zanedbává vliv samotného pixelu  $(x, y)$ .

Gradientní operátory udávající strmost obrazové funkce můžeme rozdělit do tří kategorií:

1. Operátory approximují derivace pomocí differencí, jejichž základní myšlenku jsme viděli ve vztahu (11.8). Některé operátory jsou invariantní vůči rotaci, např. Laplaceán, a mohou být počítány konvolucí s jedinou maskou. Jiné, approximující první derivaci, využívají několik masek odpovídajících příslušné orientaci. Z nich se vybere ta, která nejlépe lokálně approximuje obrazovou funkci. Výběrem jedné z masek je určen i směr gradientu (orientace).

2. Operátory založené na hledání hran v místech, kde druhá derivace obrazové funkce prochází nulou (angl. zero-crossing). Příkladem je Marrův-Hildrethové operátor a Cannyho hranový detektor, o nichž bude dále řeč.
3. Operátory snažící se lokálně approximovat obrazovou funkci poměrně jednoduchým parametrickým modelem, např. polynomem dvou proměnných.

V příští části 11.4.2 probereme jednoduché operátory approximující gradient. V části 11.4.3 vysvětlíme teorii hranové detekce podle Marra a Hildrethové. U třetí skupiny operátorů se spokojíme s předchozím principiálním vysvětlením a podrobněji je pro nedostatek místa nebudeme popisovat.

## 11.4.2 Jednoduché konvoluční masky approximující derivace obrazové funkce

Zde máme na mysli operátory, které lze vyjádřit jako masky pro konvoluci podle vztahu (11.1). Jednotlivé operátory budeme uvádět pomocí příslušného konvolučního jádra  $h$ . U směrových operátorů je tolik jader  $h$ , kolik směrů operátor rozlišuje. Z nich často uvedeme jen tolik, aby bylo možné pochopit, jak jsou jádra tvořena.

### Robertsův operátor

Nejstarší a velmi jednoduchý je Robertsův operátor [80], který používá jen okolí  $2 \times 2$  reprezentativního pixelu. Jeho konvoluční masky jsou

$$h_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad h_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (11.10)$$

Velikost gradientu se počítá podle

$$|g(x, y) - g(x+1, y+1)| + |g(x, y+1) - g(x+1, y)|. \quad (11.11)$$

Hlavní nevýhodou Robertsova operátoru je velká citlivost na šum, protože okolí použité pro approximaci je malé.

### Laplaceův operátor

Již jsme se zmínili o velmi oblíbeném Laplaceové gradientním operátoru  $\nabla^2$ , který approximuje druhou derivaci. Víme již, že je invariantní vůči otočení a udává jen velikost hran a ne její směr. V digitálním obraze je také Laplacian (11.6) approximován diskrétní konvolucí. Dvě používaná konvoluční jádra (pro 4-sousedství a 8-sousedství) v okolí  $3 \times 3$  jsou

$$h = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad h = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (11.12)$$

Někdy se používá Laplacian s větší vahou pixelů blíže reprezentativnímu bodu masky. V tomto případě se ztrácí invariantnost vůči otočení,

$$h = \begin{bmatrix} 2 & -1 & 2 \\ -1 & -4 & -1 \\ 2 & -1 & 2 \end{bmatrix}, \quad h = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix}. \quad (11.13)$$

Hlavní nevýhodou Laplaciana je velká citlivost na šum, což je ostatně při snaze approximovat druhou derivaci primitivními prostředky přirozené. Další nevýhodou jsou dvojitě odezvy na hranы odpovídající tenkým liniím v obraze. Výsledek činnosti Laplaceova operátoru jsme viděli na obr. 11.9b.

## Operátor Prewittové

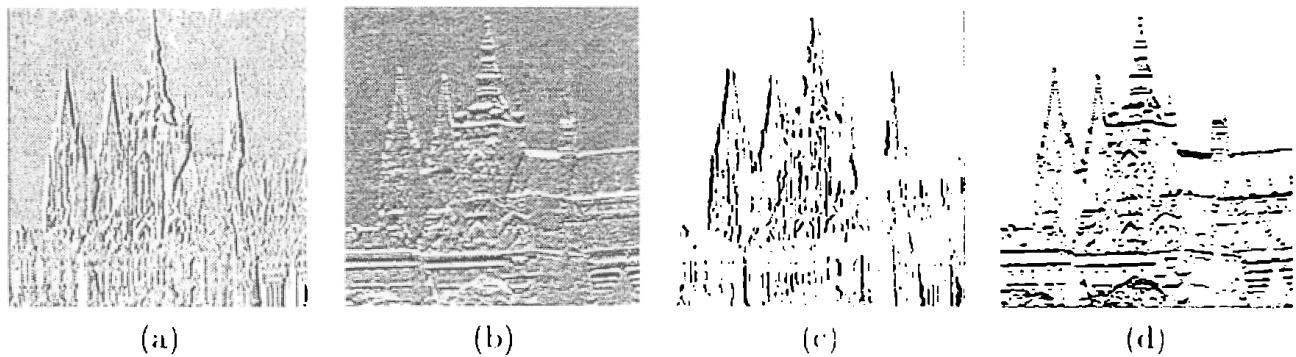
Operátor Prewittové, podobně jako Sobelův, Kirschův a Robinsonův, které ukážeme dále, approximuje první derivaci. Gradient je odhadován v okolí  $3 \times 3$  pro osm směrů. Vybrána je jedna maska z osmi, a to ta, které odpovídá největší modul gradientu. Je přirozeně možné vytvářet větší masky s podrobnějším směrovým rozlišením. Abychom ušetřili místo, budeme zde ukazovat jen první tři konvoluční masky. Ostatní si čtenář vytvoří sám pootočením.

$$h_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad h_2 = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}, \quad h_3 = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}. \quad (11.14)$$

## Sobelův operátor

$$h_1 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \quad h_2 = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}, \quad h_3 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}. \quad (11.15)$$

Sobelův operátor se často používá pro detekci vodorovných a svislých hran, na což postačí masky  $h_1, h_3$ .



**Obrázek 11.10** Detekce hran operátorem Prewittovým rozměru  $3 \times 3$ . (a) Hrany v severním směru (čím je pixel světlejší, tím je hrana silnější). (b) Hrany ve východním směru. (c) Silné hrany z (a), které byly získány prahováním. (d) Silné hrany z (b).

## Robinsonův operátor

$$h_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \quad h_2 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}, \quad h_3 = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}. \quad (11.16)$$

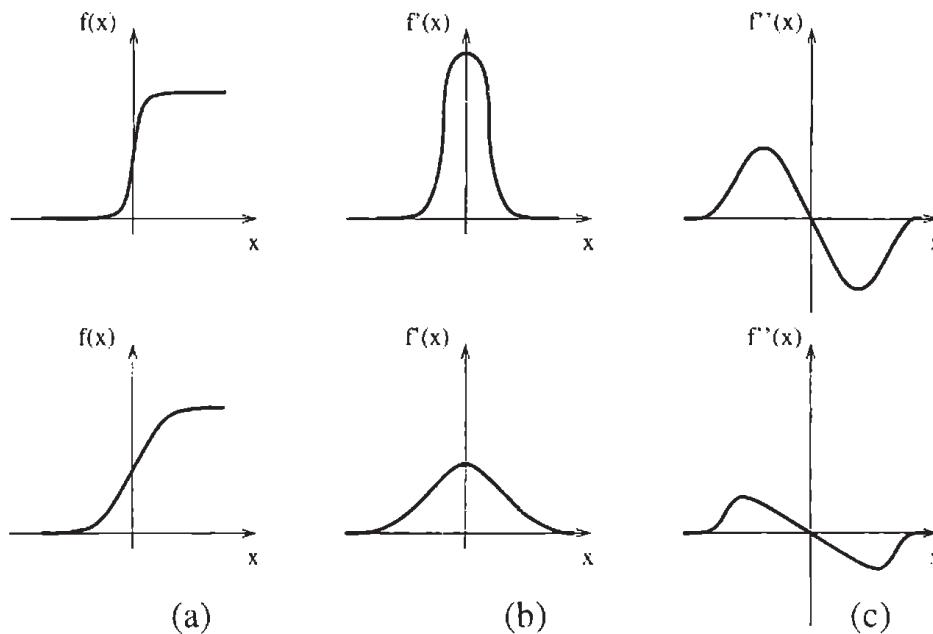
## Kirschův operátor

$$h_1 = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 0 & 3 \\ -5 & -5 & -5 \end{bmatrix}, \quad h_2 = \begin{bmatrix} 3 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & -5 & 3 \end{bmatrix}, \quad h_3 = \begin{bmatrix} -5 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & 3 & 3 \end{bmatrix}. \quad (11.17)$$

Abychom ukázali chování gradientních operátorů na příkladě, vybrali jsme operátor Prewittové. Vstupním obrázkem byl obr. 11.9a. Na obr. 11.10 je vidět, jak jsou nalezeny hranы v severním a východním směru. Poznamenejme, že světové strany používáme proto, že dobré rozliší všechny osm možných směrů.

### 11.4.3 Hrany jako průchody nulou druhé derivace obrazové funkce

Problémem výše popsaných operátorů approximujících derivaci diferencemi v malém okolí je velká závislost jejich chování na konkrétním obrázku. Velikost masky musí odpovídat velikosti detailů v obrázku. O nezávislosti na měřítku nemůže být řeč. Také citlivost na šum je značná.



**Obrázek 11.11** 1D jasový profil ilustrující polohu skokové hrany v místě průchodu 2. derivace obrazové funkce nulovou osou.

Koncem sedmdesátých let byla formulována Marrova teorie, která usilovala o matematický model detekce skokových hran odpovídající neurofyziologickým měřením na sítnici oka. Základem přístupu je hledání polohy hrany v obraze v místě průchodu druhé derivace obrazové funkce nulou. První derivace obrazové funkce nabývá svého maxima v místě hrany. Druhá derivace protíná v místě hrany nulovou hodnotu. Situaci ilustruje pro jednoduchost v 1D případě obr. 11.11, kde (a) ukazuje skokovou hranu, (b) její první derivaci a (c) průběh druhé derivace. Hledat polohu hrany v místě průchodu nulou je díky strmosti přechodu mnohem spolehlivější než na plochém maximu u první derivace.

Klíčovou otázkou je, jak robustně počítat druhou derivaci. Naše předchozí zkušenost nás nabádá k ostražitosti. Odhad druhé derivace, jak jsme už viděli v případě Laplaceánu, by měl být na šum citlivější než odhad první derivace. Uvidíme, že existuje elegantní metoda odhadu, která tuto rozumnou připomínku obejde. Klíčem k robustnímu odhadu druhé derivace je konvoluce obrazu s vyhlažujícím filtrem, na který jsou dva požadavky:

1. Filtr má být hladký a ve frekvenčním spektru přibližně odpovídající pásmové propusti, aby omezil možný počet frekvencí, při kterých k průchodu nulou může dojít.
2. Požadavek na přesnost lokalizace hrany v rovině předpokládá, že filtr bude reagovat pouze na body z blízkého okolí hrany.

Předchozí dva požadavky jsou v protikladu. Marrova teorie si všimla dobrého kompromisu (ale ne optima, jak uvidíme dále), kterým je lineární filtr, jehož koeficienty v konvoluční masce odpovídají 2D gaussovskému rozložení

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (11.18)$$

kde  $x, y$  jsou souřadnice v obrazu a  $\sigma$  je středněkvadratická odchylka, která jediným parametrem udává, na jak velkém okolí filtr operuje. Pixely blíže středu mají při filtrace větší váhu a vliv pixelů vzdálenějších než  $3\sigma$  je zanebatelný.

Naším cílem je získat druhou derivaci obrazové funkce. Konvoluci znázorníme pomocí symbolu  $*$ . Díky filtrace Gaussiánem připustíme rozmazání obrazové funkce, tj.  $G(x, y, \sigma) * f(x, y)$ . Pro odhad druhé derivace můžeme použít všeobecný Lapacián  $\nabla^2$ . Tento postup je některými autory nazýván jako *LoG operátor* (angl. Laplacian of Gaussian),  $\nabla^2[G(x, y, \sigma) * f(x, y)]$ . Díky linearitě zúčastněných operací lze zaměnit pořadí derivace a konvoluce, což je důležitý trik, a získat

$$\nabla^2(G(x, y, \sigma) * f(x, y)) = (\nabla^2 G(x, y, \sigma)) * f(x, y). \quad (11.19)$$

Hodnoty derivace Gaussiánu  $\nabla^2 G$  lze předpočítat analyticky, protože na konkrétním obrazu nezávisí! Aby se nám jednodušeji derivovalo, nahradíme  $x^2 + y^2 = r^2$ , kde  $r$  je Euklidovská vzdálenost od středu Gaussiánu. Substituce je přípustná, protože Gaussián je středově symetrický. Gaussián nyní zapíšeme jako funkci jedné proměnné  $r$ ,

$$G(r) = e^{-\frac{r^2}{2\sigma^2}}. \quad (11.20)$$

První derivace  $G'(r)$  je

$$G'(r) = -\frac{1}{\sigma^2} r e^{-\frac{r^2}{2\sigma^2}}. \quad (11.21)$$

Druhá derivace  $G''(r)$ , tj. Lapacián Gaussiánu, je

$$G''(r) = \frac{1}{\sigma^2} \left( \frac{r^2}{\sigma^2} - 1 \right) e^{-\frac{r^2}{2\sigma^2}}. \quad (11.22)$$

Když se zpětnou substitucí vrátíme k původním souřadnicím  $x, y$  a zavedeme normalizační koeficient  $c$  zajišťující, aby součet všech koeficientů v masce byl 0, dostaneme vztah poskytující hodnoty v konvoluční masce

$$h(x, y) = c \left( \frac{x^2 + y^2 - \sigma^2}{\sigma^4} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (11.23)$$

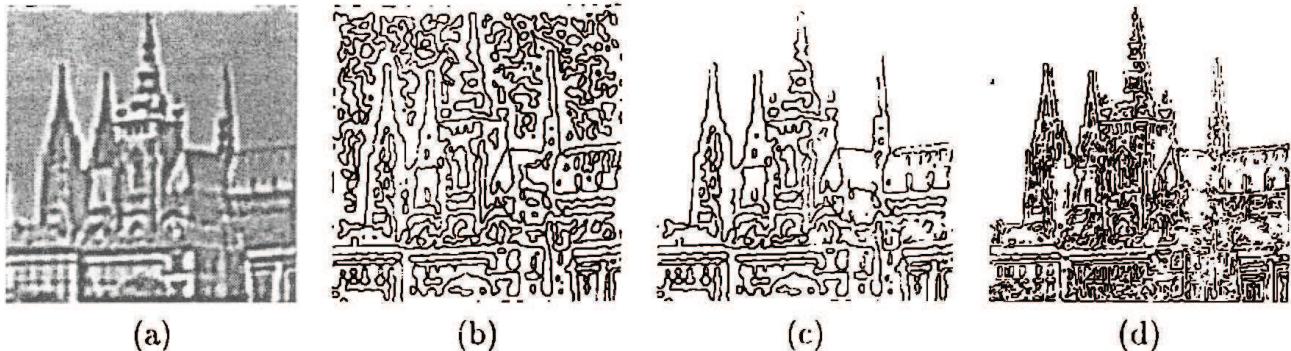
Díky svému tvaru se masce říká mexický klobouk. Ukažme si, jak vypadá approximace operátorem LoG v masce  $5 \times 5$

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}.$$

Při implementaci detektoru založeného na hledání průchodů nulou je třeba se vyhnout naivnímu řešení pomocí prahování LoG obrazu v intervalu hodnot blízkých k nule. Výsledkem by byly hodně nespojitě hrany. Lepší je použít opravdový detektor průchodů nulou, např. v masce  $2 \times 2$  s reprezentativním bodem třeba v levém horním rohu. Hrana se zde indikuje tehdy, pokud se uvnitř okna opravdu mění znaménko.

Operátor  $\nabla^2 G$ , tj. Lapacián Gaussiánů, lze efektivně approximovat pomocí diference dvou obrazů, které vznikly konvolucí s Gaussiánem o různém  $\sigma$ . Operátor bývá označován zkratkou DoG z angl. Difference of Gaussians. Volba hodnot  $\sigma$  je diskutována v [59].

Ukažme si chování různých modifikací detektorů hran opírajících se o hledání průchodu druhé derivace obrazové funkce nulou na příkladech na obr. 11.12. Jak je vidět z obrazů, mají



**Obrázek 11.12** Hledání hran jako průchodu druhé derivace obrazové funkce nulou, výchozí obrázek je výřezem z obr. 11.9a: (a) Použit operátor DoG, tj. differenze Gausiánů,  $\sigma_1 = 0,1$ ,  $\sigma_2 = 0,09$ . (b) Průchody nulou DoG obrazu. (c) Odstraněny hrany, kterým neodpovídaly významnější první derivace. (d) LoG, Laplace-Gaussián,  $\sigma = 0,2$  po odstranění hrani bez významnějších prvních derivací. Všimněte si různého měřítka hran oproti (b), (c) díky jinému  $\sigma$ .

tradiční operátory LoG a DoG a následné hledání průchodů nulou také *nevýhody*. Za prvé příliš vyhlazují ostré tvary, například ostré rohy se ztrácejí. Za druhé se snaží spojovat hrany do uzavřených křivek. Posměšně se výsledné hrany někdy označují jako "talíř špaget", i když se nekříží. Tato druhá vlastnost byla v původních pracech označována za výhodu, ale v mnoha aplikacích se ukázala spíše nevýhodou.

Tuto část jsme začali neurofyziologickou a psychologickou motivací a na podobně notě ji i ukončeme. Neurofyziologické experimenty potvrzují, že na sítnici lidského oka jsou realizovány ve shlučích buněk (nazývaných gangliony) velmi podobné operace jako  $\nabla^2 G$ . Každý ganglion reaguje na světelné podnety z kruhově symetrického okolí, které odpovídají  $-\nabla^2 G$  (tzv. excitační, protože vydávají kladnou odezvu na světelný stimul) nebo  $\nabla^2 G$  (tzv. inhibiční, protože reagují na světelný stimul záporně). Vznik Marrovy teorie měl ovšem opačnou příčinnost, k dispozici byly odezvy naměřené především na sítnicích koček a hledal se jejich matematický model.

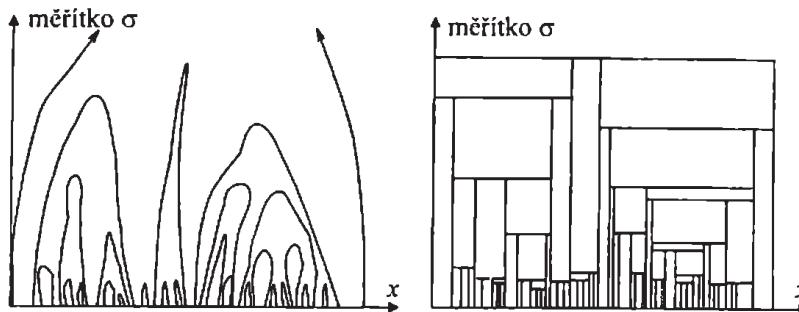
#### 11.4.4 Volba měřítka ve zpracování obrazu

Mnoho postupů počítačového vidění pracuje lokálně, z hlediska spojité matematiky na úrovni jednotlivých obrazových bodů. Příkladem je právě hledání hran. Klíčovým problémem je zde správně zvolené *měřítko*, tj. okolí použité pro výpočet. Hrany se v počítači hledají jako differenze jasu obrazových bodů v nějakém okolí. Zřídka lze přímo vybrat správnou velikost tohoto okolí. Správná velikost okolí totiž závisí na velikosti pro nás zajímavých objektů v obrazu. K tomu je ovšem nutné, abychom obraz uměli interpretovat. V době předzpracování to však většinou neumíme.

Popsanou úlohu lze řešit systémovým přístupem známým z kybernetiky pro popis složitých problémů. Zkoumaný jev se vyjádří pomocí formálního popisu (modelu) pro každé z různých rozlišení (měřítek). Potom lze studovat kvalitativní změny formálního modelu pro různá rozlišení. Takto lze získat metaznalost zkoumaného jevu, která není vidět v jediném rozlišení. Tento způsob uvažování se v počítačovém vidění velmi prosazuje.

Jak vypadají hrany nalezené v různých měřítkách jsme již viděli v části 11.4.3. Různá měřítka se vybírala změnou středně kvadratické odchyly  $\sigma$  použitého gaussovského filtru. Jevy (v našem případě hrany) v různých měřítkách nebyly vyhodnocovány automaticky, tj. lokální znalost nabytá v různých měřítkách nebyla sdružována. Tento krok nás čeká nyní.

Witkin [99] navrhl postup pro kvalitativní hodnocení signálů s ohledem na měřítko. Úloha



**Obrázek 11.13** Prostor měřítek pro případ 1D signálu. Vlevo – inflexní body druhé derivace vytvářejí křivky v prostoru  $(x, \sigma)$ . Vpravo – kvalitativní reprezentace signálu pomocí stromu intervalů, odpovídají křivkám.

byla formulována pro 1D signál, ale její zobecnění do 2D nečiní potíže. Výchozí 1D signál  $f(x)$  je vyhlazován konvolucí s 1D Gaussiánem

$$G(x, \sigma) = e^{-\frac{x^2}{2\sigma^2}}. \quad (11.24)$$

Při postupné změně středně kvadratické odchyly  $\sigma$  si lze funkci

$$F(x, \sigma) = f(x) * G(x, \sigma) \quad (11.25)$$

představit jako povrch v rovině  $(x, \sigma)$ , které se říká *prostor měřítek*. Inflexní body křivky  $F(x, \sigma_0)$  pro různé hodnoty  $\sigma_0$

$$\frac{\partial^2 F(x, \sigma_0)}{\partial x^2} = 0 \quad \frac{\partial^3 F(x, \sigma_0)}{\partial x^3} \neq 0 \quad (11.26)$$

popisují křivku  $f(x)$  kvalitativně. Polohy inflexních bodů lze nakreslit v souřadnicích  $(x, \sigma)$  jako soustavu křivek, viz obr. 11.13 vlevo. Analýzou křivek lze stanovit rozsah měřítek, při kterých nastávají zajímavé změny signálu, a to bez nutnosti obraz interpretovat. Tak lze automaticky nastavit měřítko lokálního operátoru, např. hranového detektoru.

Kvalitativní informace daná nalezenými událostmi v obraze se dá zobrazit ve tvaru jednoduchého stromu intervalů, který vyjadřuje strukturu signálu v pozorovaných měřítkách. Intervalový strom se buduje od kořene, který odpovídá největšímu měřítku  $\sigma_{max}$ . Strom se dále může prohledávat směrem k listům, tj. ve směru klesajícího  $\sigma$ . Strom intervalů se větví v bodech prostoru měřítek, kde vznikají nové křivky odpovídající inflexním bodům.

### 11.4.5 Cannyho hranový detektor

Princip různých rozlišení a hledání nejlepšího z nich realizuje Cannyho hranový detektor [4]. Základní myšlenka vychází z představy, že skokovou hranu (ve 2D obrázku si ji můžeme představit jako schod) lze hledat filtrem. Návrh tohoto filtru je formulován jako úloha variačního počtu (tj. hledání nejlepší impulsní funkce filtru) za podmínky, že budou splněny jisté požadavky na chování filtru. Vychází tedy z jiné motivace než Marrova teorie, i když byla řešena na stejném pracovišti asi o čtyři roky později. Tato úloha byla zadána doktorandovi J. Cannymu.

Detektor je optimální pro skokové hrany vzhledem ke třem kritériím:

*Detekční kritérium* požaduje, aby významné hrany nebyly přehlédnuty a aby na jednu hranu nebyly vícenásobné odezvy.

*Lokalizační kritérium* požaduje, aby rozdíl mezi skutečnou a nalezenou polohou hrany byl minimální.

Požadavek jedné odezvy zajišťuje, aby detektor ncreagoval na jednu hranu v obraze vícenásobně. Toto očekávání je již částečně zajištěno prvním kritériem. Tento požadavek je zaměřen zejména na zašuměné a nehladké hrany, což první požadavek nezajistí.

Odvození Cannynho detektora je zdlouhavé, a proto se spokojíme s rekapitulací hlavních myšlenek:

1. Nejprve byl hranový detektor formulován pro 1D signál a první dvě kritéria optimalizace. Pomocí variačního počtu s využitím programu pro symbolické odvozování bylo nalezeno explicitní řešení.
2. Po přidání třetího kritéria, tj. zabránění vícenásobným odezvám, bylo nutné hledat optimální odezvu filtru numerickou optimalizací. Výsledný filtr lze s chybou menší než 20% approximovat filtrací Gausiánem. Všimněte si podobnosti s detektorem podle Marra a Hildrethové opírající se o Laplaceán Gausiánu, viz část 11.4.3.
3. Detektor hran je zobecněn do 2D. Hrana je zde dána polohou, orientací a velikostí (představme si ji jako výšku schodu). Lze ukázat, že konvoluce s 2D Gausiánem a derivace ve směru gradientu vytváří jednoduchý a účinný diferenciální operátor, který na rozdíl od přístupu Marra a Hildrethové poskytuje i orientaci hrany.

Nechť  $G$  je 2D Gausián, viz rovnice (11.18) a předpokádejme, že chceme obrázek podrobit konvoluci s operátorem  $G_n$ , který představuje první derivaci  $G$  ve směru  $\mathbf{n}$ , tj. směru gradientu,

$$G_n = \frac{\partial G}{\partial \mathbf{n}} = \mathbf{n} \cdot \nabla G . \quad (11.27)$$

Sinér gradientu není předem znám, ale existuje jeho robustní odhad. Je-li  $f$  obrázek, normála k hraně se odhaduje jako

$$\mathbf{n} = \frac{\nabla(G * f)}{|\nabla(G * f)|} . \quad (11.28)$$

Pozice hrany odpovídá lokálnímu maximu konvoluce obrázku  $f$  s operátorem  $G_n$  ve směru  $\mathbf{n}$ .

$$\frac{\partial}{\partial \mathbf{n}} G_n * f = 0 . \quad (11.29)$$

V rovnici (11.29) dosadíme za  $G_n$  z rovnice (11.27) a získáme

$$\frac{\partial^2}{\partial \mathbf{n}^2} G * f = 0 . \quad (11.30)$$

Rovnice (11.30) ukazuje, jak najít lokální maxima ve směru kolmému na hranu. Tato operace se někdy nazývá potlačení odezv mimo maxima (angl. non-maximal suppression). Jelikož konvoluce a derivace jsou asociativní operace, může se v rovnici (11.30) nejdříve realizovat konvoluci obrázku  $f$  s Gausiánem  $G$  a potom vypočítat směrovou druhou derivaci s využitím výpočtu ve směru  $\mathbf{n}$  podle rovnice (11.28). Síla hrany (velikost gradientu intenzity  $f$ ) se vypočte podle

$$|G_n * f| = |\nabla(G * f)| . \quad (11.31)$$

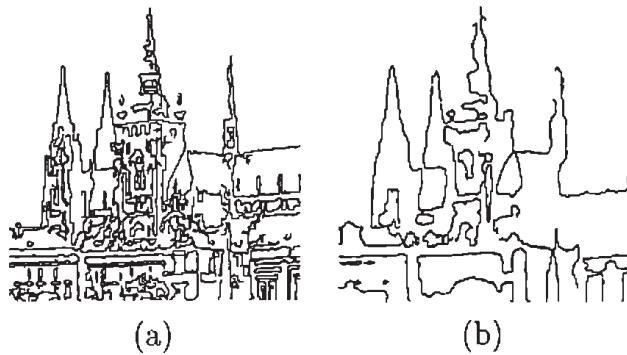
Jiný způsob zobecnění hranového detektora do dvou dimenzí byl navržen zhruba ve stejnou dobu Špačkem [90].

4. Vícenásobné odezvy na jedinou hranu způsobené šumem jsou v detekci hran běžným problémem. Výstup detektora se obvykle prahuje, a tím se zvolí, která hrana je považována za významnou. Po prahování jsou často hrany nesouvislé, ačkoliv by mnohdy měly tvořit

souvislou hranici. Tuto potíž lze potlačit *prahováním s hysterezí*. Silné hrany s modulem gradientu nad vyšším prahem jsou považovány za hranové pixely pro dané měřítko. Osamocené slabé hrany s menším modulem gradientu než vyšší práh obvykle pocházejí z šumu. Pokud jsou ale souvislé se silnou hranou, je větší naděje, že mohou být hranovými body. Za ně jsou považovány, když odpovídající modul gradientu přesahuje *nížší práh*. Vyšší a nížší práh lze nastavit automaticky podle odhadnutého poměru signálu k šumu.

5. Správné měřítko pro hranový operátor záleží na velikosti objektů v obrázku. Jelikož obraz ještě není interpretován, a tudíž se neví, co jsou objekty, je potřebné vyzkoušet všechna měřítka a sdružit informaci z nich (tyto úvahy jsme již činili v části 11.4.4). Měřítko je dánou střední kvadratickou odchylkou Gaussiánu  $\sigma$ . Detektor může mít významnější odezvy na hranu ve více měřítkách, v tom případě je použit operátor v nejmenším měřítku, protože nejlépe lokalizuje hranu.

Canny navrhl způsob *syntézy z odezv detektoru v různých měřítkách*. Nejdříve jsou označeny odezvy detektoru pro nejmenší měřítko a hrany pro větší měřítka hypotetického detektoru se syntetizují z nich. Syntetizovaná odezva se porovná se skutečnou odezvou pro příslušné větší měřítko  $\sigma$ . Hrany navíc proti syntetizovanému odhadu se zavedou jen tehdy, jsou-li silnější než předpokládala syntéza.



**Obrázek 11.14** Chování Cannyho hranového detektoru. Výchozí obrázek tvoří výřez z obr. 11.9a: (a)  $\sigma = 0,5$ .  
(b)  $\sigma = 2,8$ .

Cannyho hranový detektor představuje dnes používaný pokročilý hranový detektor. Běžné implementace většinou neobsahují zmíněnou syntézu odezv.

Obrázek 11.14a ukazuje hrany nalezené v původním obrázku 11.9a pomocí Cannyho detektoru bez syntézy odezv v různých měřítkách.

## 11.5 Obnovení obrazu při známé degradaci

### 11.5.1 Lineární model degradací

Technika předzpracování, která se snaží potlačit porušení obrazu na základě znalosti charakteru poruchy nebo jejího odhadu, se nazývá obnovením (restaurováním) [1, 43]. Zde se zaměříme na třídu lineárních degradací, pro které lze porušený obrázek  $g(x, y)$  modelovat jako konvoluci neporušeného obrázku  $f(x, y)$  s maskou  $h$ , která pokrývá celý obrázek, tj.

$$g(x, y) = \int \int_{(a,b) \in \mathcal{O}} f(a, b) h(a, b, x, y) da db + \nu(x, y),$$

kde člen  $\nu(x, y)$  popisuje aditivní šum. Většinou se předpokládá další zjednodušení, a to prostorová nezávislost filtru  $h(x, y)$ , tj. stálost příslušných koeficientů pro různé polohy v obrazu.

Potom lze degradaci vyjádřit pomocí konvoluce nebo alternativně po Fourierově transformaci jako součin Fourierova obrazu původního obrazu a Fourierova obrazu poruchy,

$$g(i, j) = (f * h)(i, j) + \nu(i, j) \quad \text{nebo} \quad G(u, v) = F(u, v) H(u, v) + N(u, v). \quad (11.32)$$

Popsaný model poruchy dobře vystihuje degradace způsobené (a) rozostřením objektivu, (b) rozmazáním pohybujícího se objektu ve scéně při dlouhých expozičních časech, (c) turbulencí atmosféry při sledování scény přes vysokou vrstvu vzduchu, např. v dálkovém průzkumu Země nebo v astronomii.

Porušení obrazu může být způsobeno i dalšími vlivy: vadami optické soustavy, nelinearitou opticko-elektrického čidla, nelinearitou nebo zrnitostí filmového materiálu, aj.

Postupy obnovení obrazu můžeme dělit na deterministické a statistické. První se hodí jen pro obrazy s malým podílem šumu. Původní obraz se zde vypočítává pomocí transformace inverzní k degradační transformaci. Příkladem deterministického obnovení je inverzní filtrace, o které budeme v této sekci ještě hovořit. Statistické metody odhadují originální obraz z obrazu zatiženého šumem o známých nebo odhadnutých statistických vlastnostech. Pro obnovení se hledá nejlepší filtr pomocí metod statistických modelů, např. metodou nejmenších čtverců. Kritériem optimality je zvolená míra podobnosti (vzdálenost) mezi původním obrazem a obrazem, který byl odhadnut metodou obnovení.

Pro výběr postupu obnovení je velmi výhodné znát vlastnosti transformace, která obraz degradovala. Je přirozené, že čím lepší je znalost degradace, tím lepší jsou i výsledky obnovení. V naprosté většině případů je naše explicitní informace o degradaci nedostačující, a proto ji musíme approximovat z existujících obrazů.

Modelování poruch dělíme podle dostupné informace na dvě skupiny - apriorní a aposteriorní. V prvním (apriorním) případě jsou parametry poruch známy nebo je lze získat před obnovením. Příkladem může být ohodnocení vlastností snímacího zařízení. Poruchy modelujeme na základě porovnání originálního a sejmutedého obrazu. Jinou možností je případ, kdy je apriorní znalost poruchy daná typem řešené úlohy. Např. při relativním pohybu kamery a objektu půjde o poruchu rozmazávání. Při modelování se určí směr a rychlosť pohybu.

Za aposteriorní označujeme znalost poruchy, kterou zjišťujeme až analýzou degradovaného obrazu. Typickým příkladem je určování charakteru poruch vyhledáváním osamělých bodů nebo přímek v obrazu a nalezením odpovídající přenosové funkce po degradaci. Jinou možností je odhadování spektrálních vlastností šumu v oblastech obrazu, o kterých víme, že jsou poměrně stejnorođé.

### 11.5.2 Matematický popis snadno odstranitelných degradací

Zde uvedeme jen tři typické a pomocí obnovení snadno odstranitelné degradace. Degradovaný obraz  $g$  může vzniknout z původního obrazu  $f$  procesem, který můžeme popsat vztahem (11.32).

#### Relativní pohyb mezi kamerou a objektem

Předpokládejme, že obrázek byl vyfotografován fotografickým přístrojem s mechanickou závěrkou. Relativní pohyb mezi kamerou a fotografovaným objektem v době otevření závěrky  $T$  způsobí rozmazání obrazu. Předpokládejme pro jednoduchost modelu konstantní rychlosť pohybu objektu  $V$  ve směru osy  $x$ . Fourierův obraz  $H(u, v)$  degradace za čas  $T$  je dán

$$H(u, v) = \frac{\sin(\pi V T u)}{\pi V u}. \quad (11.33)$$

## Rozostřený objektiv

Rozmazání objektivu špatným zaostřením tenké čočky při malé hloubce ostrosti může být popsáno jako

$$H(u, v) = \frac{J_1(a r)}{a r}, \quad (11.34)$$

kde  $J_1$  je Besselova funkce prvního druhu,  $r^2 = u^2 + v^2$ ,  $a$  je posun v obraze. Poslední parametr ukazuje, že model není prostorově invariantní.

## Turbulence atmosféry

Tuto poruchu je výhodné odstraňovat v dálkovém průzkumu Země, astronomii a vědec při sledování objektů přes silnou vrstvu vzduchu. Poruchy jsou způsobeny tepelnými nehomogenitami v atmosféře (tetelení vzduchu), které vedou k mírnému ohýbání procházejícího světla. Matematický model degradace byl stanoven pokusně [36] a zní

$$H(u, v) = e^{-c(u^2+v^2)^{\frac{5}{6}}}, \quad (11.35)$$

kde  $c$  je konstanta daná typem turbulence, která se většinou určí experimentem.

### 11.5.3 Obnovení obrazu inverzní filtrací nebo Wienerovou filtrací

Lze-li popsat porušení obrazu vztahem (11.32) a šum není vzhledem k signálu významný, potom lze úlohu řešit pomocí *inverzní filtrace*. Vztah (11.32) (jeho část vpravo ve frekvenční oblasti) po Fourierově transformaci lze při známém modelu degradace  $H$  invertovat, tj.

$$F(u, v) = G(u, v) H^{-1}(u, v) - N(u, v) H^{-1}(u, v). \quad (11.36)$$

Z rovnice (11.36) je vidět, že inverzní filtrace pracuje spolehlivě pro obrazy, které nejsou zatíženy šumem. Pokud šum není zanedbatelný, projeví se v rovnici (11.36) aditivní chyba, která se uplatňuje pro frekvence, kde má inverzní filtr malou amplitudu. To většinou nastává pro vysoké frekvence, a proto obraz obnovený inverzním filtrem má rozmažané ostré hrany. Také změny velikosti amplitudy šumu v obrazu se projeví negativně na výsledku. Navíc velikost modulu  $H(u, v)$  klesá s rostoucími frekvencemi rychleji než  $N(u, v)$ , a proto artefakty způsobené šumem mohou převážit nad užitečnou informaci. Lépe bývá použít inverzní filtraci v takovém okolí počátku roviny  $u, v$ , kde  $H(u, v)$  spolehlivě dominuje. Výsledek bývá obvykle použitelný.

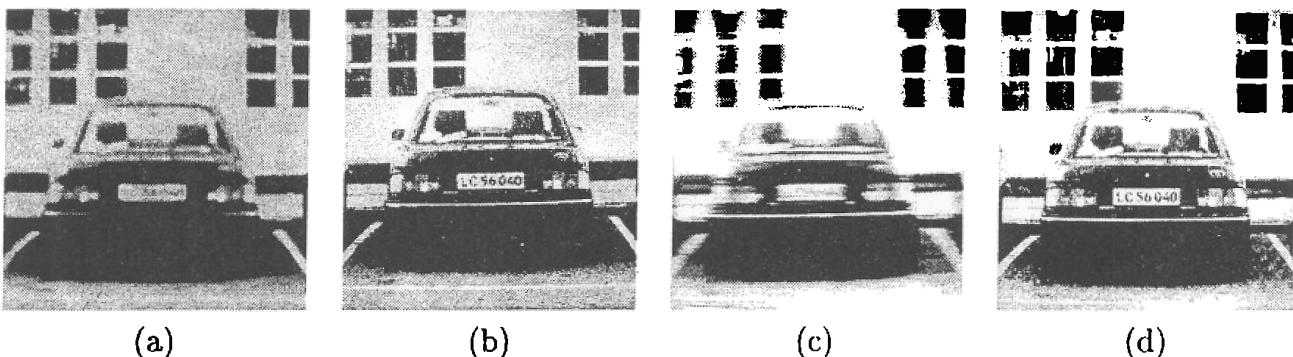
Pokud není šum zanedbatelný a dají se odhadnout jeho statistické vlastnosti, potom lze použít Wienerovu filtraci [19]. Problém obnovení obrazu je převeden na hledání původního neporušeného obrazu  $\hat{f}$  řešením přeurovené soustavy lineárních rovnic minimalizující středněkvadratickou chybu  $e^2$

$$e^2 = \mathcal{E} \left\{ (f(i, j) - \hat{f}(i, j))^2 \right\}. \quad (11.37)$$

kde  $\mathcal{E}$  označuje operátor střední hodnoty. Pokud nejsou na řešení rovnice (11.37) kladený další omezující podmínky, potom je odhad  $\hat{f}$  podmíněnou střední hodnotou ideálního obrazu  $f$  za podmínky pozorovaného obrazu  $g$ . Problémem je, že většinou není známa podmíněná pravděpodobnost správného obrazu  $f$  za podmínky, že je k dispozici pozorovaný obraz  $g$ . Optimální odhad je navíc obecně na obrazu  $g$  nelineárně závislý.

Minimalizace rovnice (11.37) je jednoduchá, když se omezíme na případ že odhad  $\hat{f}$  je lineární kombinací hodnot obrazu  $g$ . Odhad  $\hat{f}$  je potom blízký teoretickému optimu. Shoda nastane jen ve speciálním případě, když je stochastický proces popisující obrazy  $f, g$  a šum

$\nu$  stacionární a příslušné hustoty pravděpodobnosti jsou gaussovské. I když tyto podmínky nejsou vždy splněny, Wienerova filtrace se pro ně často použije, protože jiný nástroj není běžně dostupný.



**Obrázek 11.15** Restaurace dvou typických poruch Wienerovou filtrace. (a) Obraz rozmazený pohybem o 5 pixelů ve směru osy  $x$ . (b) Výsledek restaurace. (c) Obraz rozmazený díky špatně zaostřenému objektivu. (d) Výsledek restaurace. Poděkování – P. Kohout, Ředitelství policie ČR, Praha.

Označme symbolem  $H_W$  přenosovou funkci Wienerova filtrov. Pro Fourierovu transformaci odhadu neporušeného obrazu  $\hat{F}$  můžeme napsat

$$\hat{F}(u, v) = H_W(u, v) G(u, v) . \quad (11.38)$$

Neodvodíme zde, jak získat Wienerův filtr  $H_W$ , protože to lze najít např. v [71, 19, 43]. Výsledný optimální filtr je

$$H_W(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + \frac{S_{\nu\nu}(u, v)}{S_{ff}(u, v)}} , \quad (11.39)$$

kde  $H$  je Fourierova transformace degradace,  $*$  označuje komplexně sdruženou funkci,  $S_{\nu\nu}$  je výkonová spektrální hustota šumu a  $S_{ff}$  je výkonová spektrální hustota neporušeného obrazu.

Je vidět, že má-li být Wienerova filtrace použita, musí být znám charakter poruchy a statistické vlastnosti šumu. Zvláště si všimněme požadavku na znalost  $S_{ff}(u, v)$  v rovnici (11.39), což předpokládá znalost frekvenčního spektra nezkresleného obrazu. Není překvapením, že bývá těžké tuto informaci získat, když získání obrazu  $f$  je právě cílem restaurace. Je to místo na zkoušení různých spekter metodou pokusu a omylu.

Všimněte si, že ideální inverzní filtr je zvláštním případem Wienerova filtrov pro případ, že šum není znatelný, tj.  $S_{\nu\nu} = 0$ . Restaurace Wienerovou filtrace je ilustrována na obr. 11.15.

# 12. Matematická morfologie

Matematická morfologie se začala vyvíjet v šedesátých letech a svým matematickým aparátem vycházejícím z algebry nelineárních operací do značné míry při zpracování signálů či obrazů předstihuje tradiční lineární přístup, který využívá lineární kombinaci (konvoluci) bodových zdrojů představovaných Diracovými impulsy. Jde např. o předzpracování obrazu, o segmentaci s důrazem na tvar hledaných objektů, o kvantitativní popis nalezených objektů. Pro nováčka je hlavní překážkou při prvním seznámení použitá algebra, která se liší od vyučované ve standardních matematických kursech.

Hlavními protagonisty matematické morfologie byli Francouzi Matheron [60] a Serra [87], jejichž zmíněné knihy jsou matematické monografie. Klíčová pozdější díla jsou psána v podobném duchu, který je pro praktika těžko stravitelný, např. [88, 17, 13, 29].

Naším cílem je vysvětlit matematickou morfologii pro zpracování obrazů způsobem, který je pro pochopení snadný. Inspirovali jsme se u [26, 96]. Jelikož jsou nástroje matematické morfologie implementovány ve většině pokročilých balíků pro zpracování obrazu, chtěli bychom, aby se je čtenář naučil kvalifikovaně používat. Operátory matematické morfologie se obvykle používají tam, kde je požadavek na krátký čas zpracování. Aplikačními oblastmi jsou biologie, materiálový výzkum, geologie, kriminalistika, obrazová inspekce v průmyslu, rozpoznávání znaků a dokumentů, aj. Je přirozené, že morfologické metody lze použít nejen pro 2D obrazy, ale lze jimi vyjádřit i zpracování 1D signálů [56, 57].

## 12.1 Základní morfologické pojmy

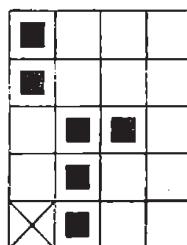
Matematická morfologie využívá vlastností bodových množin, výsledky z integrální geometrie a topologie. Výchozím předpokladem je představa, že reálné obrázky lze modelovat pomocí *bodových množin libovolné dimenze* (např.  $N$ -rozměrný euklidovský prostor). Dvojrozměrný euklidovský prostor  $\mathcal{E}^2$  a systém jeho podmnožin je přirozeným definičním oborem pro popis rovinných útváří. Předpokládáme zde, že čtenář rozumí množinovým pojmem jako podmnožina ( $\subset$  or  $\supset$ ), průnik ( $\cap$ ), sjednocení ( $\cup$ ), prázdná množina  $\emptyset$  a množinový doplněk ( $^c$ ). Množinový rozdíl je definován jako

$$X \setminus Y = X \cap Y^c. \quad (12.1)$$

V počítačovém vidění se používá digitální protějšek euklidovského prostoru. Pro binární matematickou morfologii zpracovávající dvojúrovňový obrázek je základním popisným kamenem množina dvojic celých čísel ( $\in \mathbb{Z}^2$ ). Pro šestotónovou matematickou morfologii zpracovávající obrázek s více úrovněmi jasu jsou základem popisu trojice ( $\in \mathbb{Z}^3$ ).

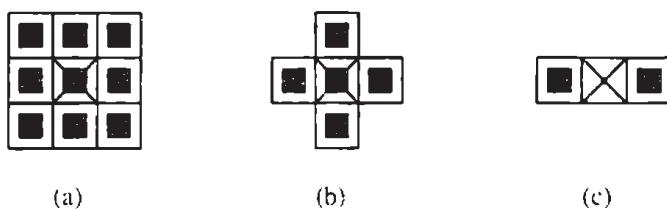
Začneme binárními obrazy, na které lze nahlížet jako na podmnožinu 2D prostoru všech celých čísel,  $\mathbb{Z}^2$ . Bod (pixel) je reprezentován dvojicí celých čísel vzhledem ke dvěma souřadným osám diskrétní mřížky. Jednotková délka podél os odpovídá periodě vzorkování. O diskrétním rastru mluvíme, když je diskrétní mřížka obohacena o relaci sousedství. Tato představa se hodí jak pro čtvercové tak i hexagonální mřížky. V tomto skriptu se předpokládá čtvercový rastr.

Binární obraz lze vyjádřit jako 2D bodovou množinu. Body objektů v obrazu reprezentují množinu  $X$ , což odpovídá pixelům s hodnotou jedna. Body doplňku  $X^c$  popisují pozadí a reprezentují se pixely s hodnotou nula. Počátek (označován v našich příkladech křížkem ve směru diagonál) má souřadnice  $(0, 0)$  a souřadnice ostatních bodů  $(x, y)$  mají stejný význam, jako je v matematice obvyklé. Obr. 12.1 ukazuje příklad bodové množiny. Body objektu jsou zapsány jako malé černé čtverce. Každý bod  $x$  diskrétního obrazu, např.  $X = \{(1, 0), (1, 1), (1, 2), (2, 2), (0, 3), (0, 4)\}$  se může chápat jako radiusvektor vzhledem k počátku  $(0, 0)$ .



Obrázek 12.1 Příklad bodové množiny.

*Morfologická transformace*  $\Psi$  je dána relací mezi obrazem (bodová množina  $X$ ) s jinou, typicky menší bodovou množinou  $B$ , které se říká *strukturní element*. Strukturní element  $B$  je vztažen k "lokálnímu" počátku  $\mathcal{O}$ , kterému se říká reprezentativní bod. Některé typické strukturní elementy jsou na obr. 12.2. Obrázek 12.2c ukazuje nepříliš žádoucí případ, kdy reprezentativní bod  $\mathcal{O}$  není bodem strukturního elementu  $B$ .



Obrázek 12.2 Typické strukturní elementy.

Aplikaci morfologické transformace  $\Psi(X)$  na obraz  $X$  si představíme, jako bychom strukturní element  $B$  systematicky posouvali po obrazu. Uvažujme, že  $B$  je umístěn v nějaké poloze v obrazu. Výsledek relace mezi obrazem  $X$  a strukturním elementem  $B$  se zapíše do výstupního obrazu v reprezentativním pixelu. Pro nyní probírané binární obrazy je výsledek relace buď 0 nebo 1.

*Dualita* morfologických operací vyplývá z množinového doplňku. Ke každé morfologické transformaci  $\Psi(X)$  existuje duální transformace  $\Psi^*(X)$ ,

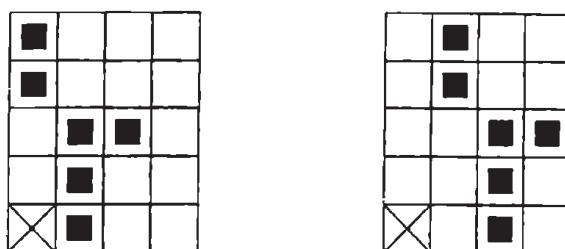
$$\Psi(X) = (\Psi^*(X^c))^c . \quad (12.2)$$

*Translace* bodové množiny  $X$  o radiusvektor  $h$  se označuje  $X_h$  a je definována vztahem

$$X_h = \{p \in \mathcal{E}^2, p = x + h \text{ pro některá } x \in X\} . \quad (12.3)$$

Translace je ilustrována na obr. 12.3.

Základními transformacemi matematické morfologie jsou dilatace, eroze, otevření a uzavření. Postupně si uvedeme jejich definice a vlastnosti.

Obrázek 12.3 Translace o vektor  $\{(1,0)\}$ .

## 12.2 Binární matematická morfologie

Pro vyjádření morfologických transformací pro binární obrázky použijeme Minkowského formalismus [26], protože je bližší k výkladu v obvyklých matematických kurzech. Alternativou by byl Serrův formalismus [87].

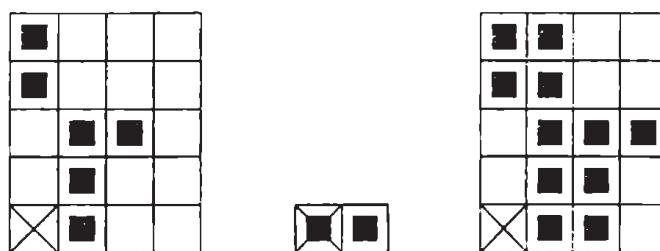
### 12.2.1 Dilatace

*Dilatace*  $\oplus$  skládá body dvou množin pomocí vektorového součtu (též Minkowského množinový součet, např.  $(a, b) + (c, d) = (a+c, b+d)$ ). Dilatace  $X \oplus B$  je bodovou množinou všech možných vektorových součtů pro dvojice pixelů, vždy pro jeden z množiny  $X$  a jeden z množiny  $B$

$$X \oplus B = \{p \in \mathcal{E}^2 : p = x + b, \quad x \in X, b \in B\} . \quad (12.4)$$

Obr. 12.4 ukazuje dilataci na jednoduchém příkladě.

$$\begin{aligned} X &= \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4)\} \\ B &= \{(0,0), (1,0)\} \\ X \oplus B &= \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4), \\ &\quad (2,0), (2,1), (2,2), (3,2), (1,3), (1,4)\} \end{aligned}$$



Obrázek 12.4 Dilatace.

Obr. 12.5 ukazuje dilataci pro reálný obraz. Vstupním obrazem (vlevo) je znak Českého vysokého učení technického rozměru  $256 \times 256$ . Je použit strukturní element o rozměru  $3 \times 3$ , viz obr. 12.2a. Výsledek dilatace je na obr. 12.5 vpravo. V tomto případě jsme použili *isotropický* strukturní element, což znamená, že se morfologická transformace chová stejně ve všech směrech, tj. objekty expandují. Vidíme, že objekty se rozrostly o jednu „slupku“ na úkor pozadí. Také díry, které měly tloušťku jeden bod, se zaplnily.

Dilataci s isotropickým strukturním elementem  $3 \times 3$  si můžeme představit jako transformaci, která změní všechny body pozadí sousedící s objekty na body objektů.



**Obrázek 12.5** Dilatace jako isotropická expanze.

Uvedeme bez důkazů několik zajímavých vlastností dilatace, detaily lze najít v monografii [87] nebo úvodním článku [28].

Dilatace je komutativní operace

$$X \oplus B = B \oplus X . \quad (12.5)$$

Je také asociativní

$$X \oplus (B \oplus D) = (X \oplus B) \oplus D . \quad (12.6)$$

Dilataci můžeme vyjádřit jako sjednocení posunutých bodových množin

$$X \oplus B = \bigcup_{b \in B} X_b . \quad (12.7)$$

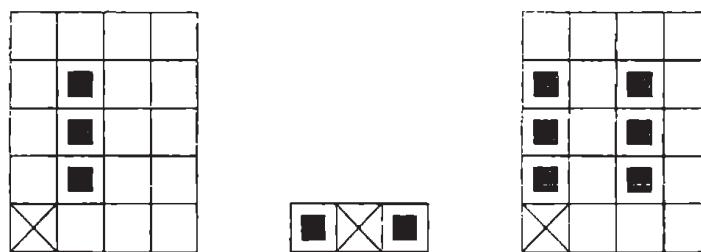
Dilatace je invariantní vůči posunu

$$X_h \oplus B = (X \oplus B)_h . \quad (12.8)$$

Rovnice (12.7) a (12.8) ukazují význam posunu pro rychlou implementaci dilatace. Platí to obecněji pro operace binární morfologie implementované na sériových počítačích, kde se v jednom počítačovém slově reprezentuje několik pixelů binárního obrazu.

Dilatace je rostoucí transformace.

$$\text{Je-li } X \subseteq Y , \text{ potom } X \oplus B \subseteq Y \oplus B . \quad (12.9)$$



**Obrázek 12.6** Dilatace pro případ, kdy počátek není prvkem strukturního elementu.

Dilatace se používá samostatně k zaplnění malých dér, úzkých zálivů a jako stavební kámen složitějších operací. Dilatace zvětšuje objekty. Má-li se zachovat jejich původní rozměr, kombinuje se dilatace s erozí, která bude popsána v následujícím odstavci.

Obr. 12.6 ukazuje výsledek dilatace v patologickém případě, když reprezentativní bod není prvkem strukturního elementu. Výsledek se podstatně liší od vstupního obrazu, protože byla porušena souvislost.

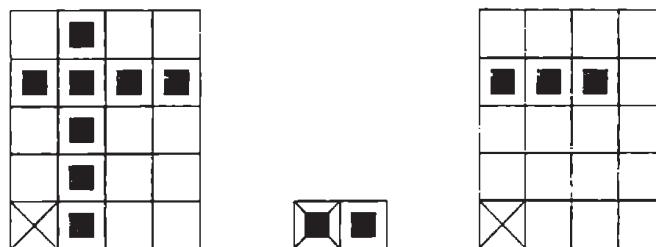
## 12.2.2 Eroze

*Eroze* je duální operací k dilataci. Ani dilatace ani eroze nejsou invertovatelné. Eroze  $\ominus$  skládá dvě množiny podle předpisu

$$X \ominus B = \{p \in \mathcal{E}^2 : p + b \in X \text{ pro každé } b \in B\}. \quad (12.10)$$

Předchozí vztah říká, že pro každý bod obrazu  $p$  se ověřuje, zda pro všechna možná  $p + b$  leží výsledek v  $X$ . Pokud ano, zapíše se v reprezentativním bodě do výsledného obrázku 1 a v opačném případě 0. Obr. 12.7 ukazuje bodovou množinu  $X$  erodovanou strukturním elementem  $B$ , který není všesměrový. Svislá čára zmizí.

$$\begin{aligned} X &= \{(1,0), (1,1), (1,2), (0,3), (1,3), (2,3), (3,3), (1,4)\} \\ B &= \{(0,0), (1,0)\} \\ X \ominus B &= \{(0,3), (1,3), (2,3)\} \end{aligned}$$



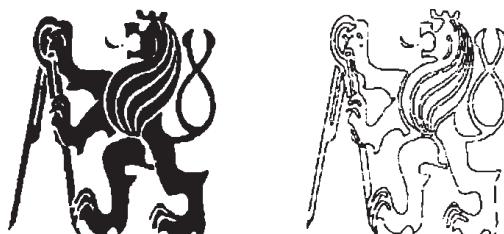
Obrázek 12.7 Eroze.

Obr. 12.8 demostruje erozi isotropickým strukturním elementem  $3 \times 3$  z obr. 12.2a na stejném vstupním obrázku jako obr. 12.5. Ve výsledku zmizely čáry tloušťky 1.



Obrázek 12.8 Eroze jako isotropické smrštění objektů.

Nyní už bychom uměli snadno najít obrys objektů, a to odečtením erodovaného obrázku od původního obrázku, viz obr. 12.9.



Obrázek 12.9 Obrys (vpravo) jako rozdíl původního (vlevo) a erodovaného obrázku.

Eroze se používá pro zjednodušení struktury objektů – objekty tloušťky 1 se ztratí, a tak se složitější objekt rozdělí na několik jednodušších.

Existuje ekvivalentní vyjádření eroze [61]. Připomeňme, že  $B_p$  označuje množinu  $B$  posunutou o vektor  $p$

$$X \ominus B = \{p \in \mathcal{E}^2 : B_p \subseteq X\}. \quad (12.11)$$

Erozi můžeme také interpretovat jako systematické posouvání strukturního elementu  $B$  po obrazu  $X$ . Když je  $B$  posunutý o vektor  $p$  obsažen v obrazu  $X$ , potom bod odpovídající reprezentativnímu bodu  $B$  patří do eroze  $X \ominus B$ .

Implementace eroze se může zjednodušit, když si všimneme, že erozi obrazu  $X$  strukturním elememtem  $B$  můžeme vyjádřit jako průnik všech posunů obrazu  $X$  o vektory  $-b \in B$

$$X \ominus B = \bigcap_{b \in B} X_{-b}. \quad (12.12)$$

Poznamenejme, že se zde používaná definice eroze liší od [87]. Tam  $\ominus$  označuje Minkowského rozdíl. Ve vyjádření průniku posunutých strukturních elementů by byly použity vektory  $b \in B$ . Náš případ se liší přidaným záporným znaménkem.

Pokud je reprezentativní bod prvkem strukturního elementu, potom je eroze antiextenzivní transformací, tj. když  $(0,0) \in B$ , potom  $X \ominus B \subseteq X$ . Eroze je také invariantní vůči posunu

$$X_h \ominus B = (X \ominus B)_h, \quad (12.13)$$

$$X \ominus B_h = (X \ominus B)_{-h} \quad (12.14)$$

a podobně jako dilatace je rostoucí transformaci.

$$\text{Je-li } X \subseteq Y, \text{ potom } X \ominus B \subseteq Y \ominus B. \quad (12.15)$$

Nechť jsou  $B, D$  strukturními elementy a  $D \subseteq B$ . Potom je eroze pomocí  $B$  “agresivnější než pomocí”  $D$ ; tj. když  $D \subseteq B$ , potom  $X \ominus B \subseteq X \ominus D$ . Tato vlastnost umožňuje uspořádat eroze podle příslušných strukturních elementů podobného tvaru, ale různého rozměru.

Nechť  $\check{B}$  označuje *symetrickou množinu* (někdy se také říká *transponovaná množina*) vůči reprezentativnímu bodu  $O$

$$\check{B} = \{-b : b \in B\}. \quad (12.16)$$

Například

$$\begin{aligned} B &= \{(1, 2), (2, 3)\} \\ \check{B} &= \{(-1, -2), (-2, -3)\} \end{aligned} \quad (12.17)$$

Již jsme se zmínili, že eroze a dilatace jsou duální transformace,

$$(X \ominus Y)^C = X^C \oplus \check{Y}. \quad (12.18)$$

Eroze se také od dilatace svými vlastnostmi liší. Eroze není na rozdíl od dilatace komutativní

$$X \ominus B \neq B \ominus X. \quad (12.19)$$

Pro erozi kombinovanou s průnikem platí

$$\begin{aligned} (X \cap Y) \ominus B &= (X \ominus B) \cap (Y \ominus B), \\ B \ominus (X \cap Y) &\supseteq (B \ominus X) \cup (B \ominus Y). \end{aligned} \quad (12.20)$$

Na druhou stranu nelze zaměnit pořadí dilatace a průniku. Dilatace průniku dvou obrazů je obsažena v průniku jejich dilatací,

$$(X \cap Y) \oplus B = B \oplus (X \cap Y) \subseteq (X \oplus B) \cap (Y \oplus B).$$

Pořadí eroze a množinového sjednocení zaměňovat lze, což umožnuje rozložit složitější strukturní elementy na sjednocení jednodušších

$$\begin{aligned} B \oplus (X \cup Y) &= (X \cup Y) \oplus B = (X \oplus B) \cup (Y \oplus B), \\ (X \cup Y) \ominus B &\supseteq (X \ominus B) \cup (Y \ominus B), \\ B \ominus (X \cup Y) &= (X \ominus B) \cap (Y \ominus B). \end{aligned} \quad (12.21)$$

Postupná dilatace (resp. eroze) obrazu  $X$  nejdříve strukturním elementem  $B$  a potom strukturním elementem  $D$  je totožná jako dilatace (resp. eroze) obrazu  $X$  pomocí  $B \oplus D$ ,

$$\begin{aligned} (X \oplus B) \oplus D &= X \oplus (B \oplus D), \\ (X \ominus B) \ominus D &= X \ominus (B \oplus D). \end{aligned}$$

### 12.2.3 Transformace tref či miň

Transformace tref či miň (angl. hit or miss) je morfologický operátor, který indikuje shodu strukturního elementu a části obrazu. Strukturní element představuje vzor, který se vyhledává, a může například sloužit k vyhledávání rohů, hranic objektů, pro ztenčování, což vysvětlíme dále v části 12.4.3.

Dosud popisované operace využívaly strukturní element  $B$ , kterým byl testován výskyt bodů v obrazu  $X$ . Nyní budeme potřebovat testovat i to, zda nějaké body do  $X$  nepatří. K tomu slouží *složený strukturní element*, což je dvojice disjunktních množin  $B = (B_1, B_2)$ . Transformace *tref či miň* označovaná  $\otimes$  je definována jako

$$X \otimes B = \{x : B_1 \subset X \text{ a } B_2 \subset X^c\}. \quad (12.22)$$

Má-li být bod  $x$  ve výsledné množině, musí být současně splněny dvě podmínky. Za prvé musí být část  $B_1$  složeného strukturního elementu s reprezentativním bodem v poloze  $x$  obsažena v  $X$ . Za druhé nesmí být část  $B_2$  složeného strukturního elementu obsažena v  $X^c$ .

Transformace tref či miň se tedy chová jako ověřování lokální shody mezi částí obrazu  $X$  a strukturním elementem  $(B_1, B_2)$ . Transformaci tref či miň můžeme také vyjádřit pomocí erozí a dilatací

$$X \otimes B = (X \ominus B_1) \cap (X^c \ominus B_2) = (X \ominus B_1) \setminus (X \oplus \check{B}_2). \quad (12.23)$$

### 12.2.4 Otevření a uzavření

Víme již, že dilatace a eroze nejsou navzájem inverzní zobrazení. Jejich kombinace jsou dalšími významnými morfologickými transformacemi - otevřením a uzavřením. Výsledkem obou je zjednodušený obraz, který obsahuje inéně detailů.

Eroze následovaná dilatací vytváří novou transformaci zvanou *otevření*. Otevření množiny  $X$  strukturním elementem  $B$  se označuje  $X \circ B$  a je definováno jako

$$X \circ B = (X \ominus B) \oplus B. \quad (12.24)$$

Dilatace následovaná erozí je uzavřením. *Uzavření* množiny  $X$  strukturním elementem  $B$  se označuje  $X \bullet B$  a je definováno jako

$$X \bullet B = (X \oplus B) \ominus B. \quad (12.25)$$

Pokud se obraz  $X$  nezmění po otevření strukturním elementem  $B$ , říkáme, že je obraz  $X$  otevřený vzhledem k  $B$ . Podobně, pokud se obraz  $X$  nezmění po uzavření strukturním elementem  $B$ , říkáme, že je uzavřený vzhledem k  $B$ .



**Obrázek 12.10** Otevření (výchozí obraz je vlevo).



**Obrázek 12.11** Uzavření (výchozí obraz je vlevo).

Otevření a uzavření izotropickým strukturním elementem se používá pro odstranění detailů v obraze, které jsou menší než strukturní element. Celkový tvar objektu se tak neporuší. Otevření oddělí objekty spojené úzkou šíjí a tak zjednoduší strukturu objektů, což je vidět z obr. 12.10.

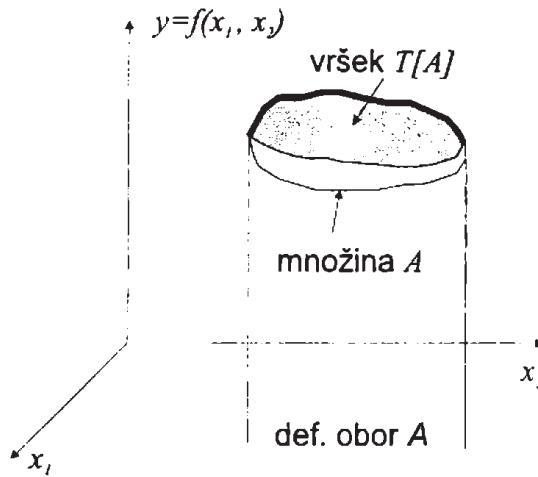
Uzavření spojí objekty, které jsou blízko u sebe, zaplní malé díry a vyhladí obrys tím, že zaplní úzké zálivy. „Malý“, „blízký“ a „úzký“ je relativní vzhledem k velikosti strukturního elementu. Vlastnosti uzavření jsou vidět na obr. 12.11.

Na rozdíl od eroze a dilatace je otevření a uzavření invariantní vzhledem k posunu strukturního elementu.

Z rovnic (12.9), (12.15) vyplývá, že jak otevření tak i uzavření jsou rostoucí transformace. Otevření je antieextenzivní ( $X \circ B \subseteq X$ ) a uzavření je extenzivní ( $X \subseteq X \bullet B$ ).

Otevření a uzavření, podobně jako dilatace a eroze, jsou duálními transformacemi

$$(X \bullet B)^C = X^C \circ \bar{B}. \quad (12.26)$$



Obrázek 12.12 Vršek množiny  $A$  odpovídá maximálním hodnotám funkce  $f(x_1, x_2)$ .

Další důležitou vlastností otevření a uzavření je idempotentnost, což znamená, že opakování použití těchto operací nemění předchozí výsledek. Proto má smysl mluvit o otevřené či uzavřené množině vzhledem ke strukturálnímu elementu  $B$ . Formálně

$$X \circ B = (X \circ B) \circ B, \quad X \bullet B = (X \bullet B) \bullet B. \quad (12.27)$$

## 12.3 Šedotónová matematická morfologie

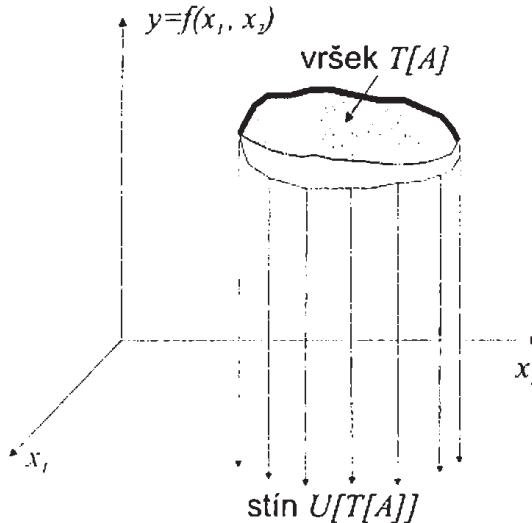
Operace matematické morfologie zavedené pro binární obrázky lze snadno zobecnit pro obrazy s více jasovými úrovněmi. Klíčovou roli zde hrají operace *min* a *max*. Eroze (resp. dilatace) obrazu přiřazuje každému pixelu minimální (resp. maximální) hodnotu v okolí okamžitého bodu vstupního obrazu. Také strukturní element je bohatší než u binárních obrazů, kde určoval pouze uvažované okolí. V případě šedotónových obrazů je strukturní element funkcí dvou proměnných a ovlivňuje, jakým způsobem se berou v úvahu hodnoty obrazu v okolí. Hodnota strukturního elementu je přičtena (nebo odečtena), když se v okolí počítá maximum (nebo minimum).

Toto rozšíření morfologie si lze představit pomocí *topografického pohledu* na šedotónové obrazy. Hodnota jasu může být interpretována jako výška v hypotetické krajině. Světlá a tmavá místa v obraze odpovídají kopcem a prohlubním. Morfologický přístup umožňuje najít globální vlastnosti obrazu, které v krajině odpovídají topograficky významným jevům, jako údolí, rozvodí, hřebeny hor, atd.

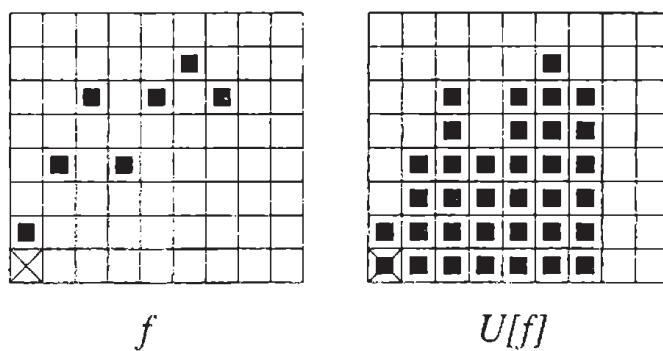
V našem výkladu jsme se nechali inspirovat [26] pro vysvětlení šedotónové dilatace a eroze, pro což zavedeme pojmy *stín* (angl. *umbra*) a *vršek* (angl. *top of the surface*) bodové množiny. Šedotónovou dilataci vyjádříme jako dilataci stínů.

### 12.3.1 Vršek, stín, šedotónová dilatace a eroze

Uvažujme bodovou množinu  $A$  v  $n$ -rozměrném euklidovském prostoru,  $A \subset \mathcal{E}^n$ . Předpokládejme dále, že prvních  $(n-1)$  souřadnic množiny tvoří definiční obor a  $n$ -tá souřadnice odpovídá funkční hodnotě funkce nebo funkci v bodě (pro šedotónové obrazy  $n = 3$ ). Tato interpretace bodové množiny odpovídá již zmíněnému topografickému pohledu na 3D euklidovský prostor, kde body jsou dány trojicemi souřadnic, první dvě z nich jsou souřadnice v rovině (definiční obor) a třetí udává výšku (funkční hodnota, v obrazech jas).



Obrázek 12.13 Stín množiny je celý prostor pod vrškem.



Obrázek 12.14 Příklad 1D funkce (vlevo) a jejího stínu (vpravo).

*Vršek* množiny  $A$  je funkce definovaná na  $(n - 1)$ -rozměrném definičním oboru. Pro každou  $(n - 1)$ -tici je vršek nejvyšší hodnota zbylé poslední souřadnice množiny  $A$ , jak je ilustrována na obr. 12.12. Pro euklidovský prostor má nejvyšší hodnota význam suprema.

Nechť  $A \subseteq \mathcal{E}^n$  a nechť definiční obor  $F = \{x \in \mathcal{E}^{n-1} \text{ pro některá } y \in \mathcal{E}, (x, y) \in A\}$ . *Vršek* množiny  $A$ , označovaný  $T[A]$ , je zobrazením  $F \rightarrow \mathcal{E}$  definovaným jako

$$T[A](x) = \max\{y, (x, y) \in A\}. \quad (12.28)$$

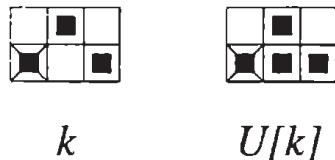
Dalším pojmem, který potřebujeme zavést, je *stín* funkce  $f$  definovaný na  $(n - 1)$ -dimenzionální podmnožině (definičním oboru) množiny  $A$ . Stín si představujeme jako oblast stínu vrženého neprůhledným objektem za předpokladu velmi vzdáleného bodového zdroje světla v nadhlavníku. V matematické morfologii je stínem funkce  $f$  množina sestávající z vršku  $f$  a celého prostoru pod ním, viz obr. 12.13.

Nyní zavedeme stín a vršek formálně. Nechť  $F \subseteq \mathcal{E}^{n-1}$  a  $f: F \rightarrow \mathcal{E}$ . *Stín* funkce  $f$  se označuje  $U[f]$ ,  $U[f] \subseteq F \times \mathcal{E}$  a je definovaný vztahem

$$U[f] = \{(x, y) \in F \times \mathcal{E}, y \leq f(x)\}. \quad (12.29)$$

Je vidět, že stínem stínu funkce  $f$  je opět stín.

Ilustrujme pojmy stín a vršek pro případ 1D signálu (můžeme ho chápout jako 1D šedotónový obraz). Obr. 12.14 ukazuje funkci  $f$ , která může být vrškem množiny a její stín.



**Obrázek 12.15** Strukturní element; 1D funkce (vlevo) a její stín (vpravo).

$U[f] \oplus U[k]$        $T[U[f] \oplus U[k]] = f \oplus k$

**Obrázek 12.16** 1D příklad šedotónové dilatace. Stíny 1D funkce  $f$  a strukturního elementu  $k$  jsou nejdříve dilatovány,  $U[f] \oplus U[k]$ . Výsledkem je vršek této dilatované množiny  $f \oplus k = T[U[f] \oplus U[k]]$ .

Nyní již můžeme definovat šedotónovou dilataci jako vršek dilatace jejich stínů. Nechť  $F, K \subseteq \mathcal{E}^{n-1}$ ,  $f: F \rightarrow \mathcal{E}$  a  $k: K \rightarrow \mathcal{E}$ . Šedotónová dilatace  $\oplus$  funkce  $f$  s funkcí  $k$ ,  $f \oplus k: F \oplus K \rightarrow \mathcal{E}$  je definována jako

$$f \oplus k = T \{U[f] \oplus U[k]\}. \quad (12.30)$$

Všimněte si, že v předchozí rovnici je  $\oplus$  na levé straně dilatací v prostoru šedotónových obrazů a  $\oplus$  na pravé straně je dilatací binárních obrazů. Nezaváděli jsme nový symbol, protože zde nehrozí nebezpečí zmatení. Totéž platí i pro erozi  $\ominus$ , kterou teprve zavedeme.

Podobně jako u binární dilatace představuje jedna funkce  $f$  obraz a druhá  $k$  malý strukturní element. Na obr. 12.15 je vidět diskretizovaná funkce  $k$ , která je strukturním elementem. Obr. 12.16 znázorňuje dilataci stínu funkce  $f$  (z obr. 12.14) stínem funkce  $k$ .

Tato definice sice názorně vysvětluje, co je šedotónová dilatace, ale neposkytuje rozumný algoritmus pro výpočet. Uvidíme, že výpočetně efektivnější postup vede přes maximum součtu v množině

$$(f \oplus k)(x) = \max \{f(x - z) + k(z), z \in K, x - z \in F\}. \quad (12.31)$$

Výpočetní složitost je stejná jako při konvolučních operacích v lineární filtrace, kde se sčítají součiny.

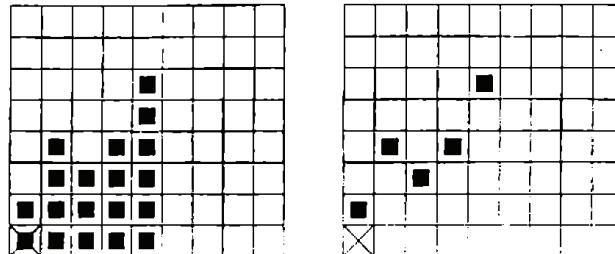
Definice šedotónové eroze je analogická k dilataci. Šedotónová eroze dvou funkcí (bodových množin): (1) Najde jejich stíny. (2) Eroduje je binární erozí. (3) Výpočte výsledek jako vršek množiny.

Nyní zavedeme šedotónovou erozi formálně. Nechť  $F, K \subseteq \mathcal{E}^{n-1}$ ,  $f: F \rightarrow \mathcal{E}$  a  $k: K \rightarrow \mathcal{E}$ . Eroze  $\ominus$  množiny  $f$  množinou  $k$ ,  $f \ominus k: F \ominus K \rightarrow \mathcal{E}$  je definována jako

$$f \ominus k = T \{U[f] \ominus U[k]\}. \quad (12.32)$$

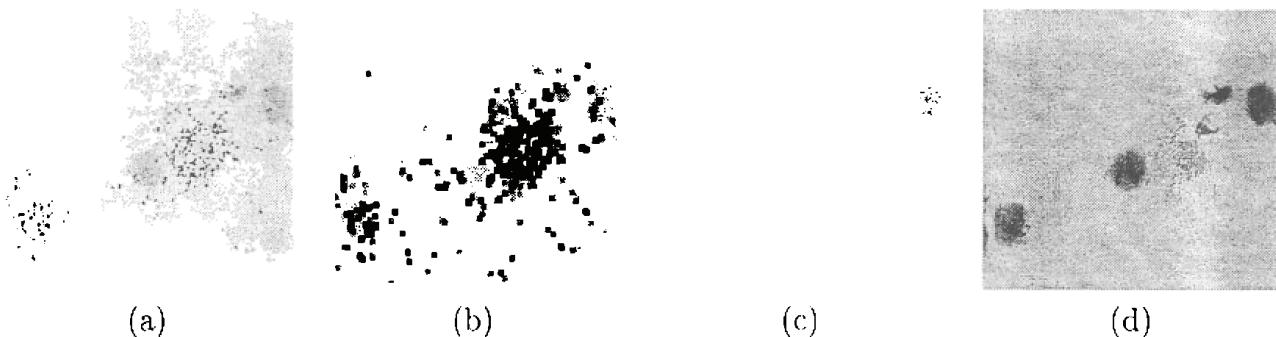
Aby se snížila výpočetní náročnost, probíhá skutečný výpočet jinak, jako minimum množiny diferencí (všimněte si podobnosti s korelací):

$$(f \ominus k)(x) = \min_{z \in K} \{f(x + z) - k(z)\}. \quad (12.33)$$



$$U[f] \oplus U[k] \quad T[U[f] \oplus U[k]] = f \oplus k$$

**Obrázek 12.17** 1D příklad šedotónové eroze. Stíny 1D funkce  $f$  a strukturního elementu  $k$  se nejdříve erodují,  $U[f] \ominus U[k]$ . Vršek poskytne výsledek,  $f \ominus k = T[U[f] \ominus U[k]]$ .



**Obrázek 12.18** Příklad morfologického předzpracování. (a) Buňky v mikroskopickém obrazu zatížené šumem. (b) Erovaný obrázek. (c) Dilatace obrazu b, šum zmizel. (d) Rekonstruované buňky. Poděkování. P. Kodl, Rockwell Automation Research Center, Praha.

Ukažme si předzpracování metodami matematické morfologie na příkladu zpracování mikroskopických obrazů buněk zatížených šumem, viz 12.18a. Cílem je omezit šum a najít jednotlivé buňky. Obr. 12.18b ukazuje erozi výchozího obrazu a obr. 12.18c ilustruje dilataci původního obrazu. V obou případech byl použit strukturní element  $3 \times 3$ . Po této transformaci šum výrazně poklesl. Rekonstruované buňky se najdou rekonstrukční operací, která bude popsána v části 12.4.4. Zde je výchozí obraz použit jako maska a dilatovaný obraz 12.18c je vstupem rekonstrukce. Výsledek je ukázán na obr. 12.18d, ve kterém tmavé skvrny označují nalezené buňky.

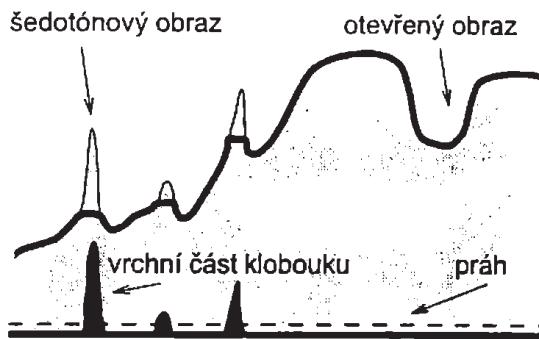
### 12.3.2 Věta o homeomorfismu stínu, vlastnosti eroze, dilatace, otevření a uzavření

Operace vršek množiny je vždy levou inverzí operace stínů. Nejsilnějším tvrzením, které lze vyvodit z řečeného, je fakt, že stín vršku bodové množiny  $A$  obsahuje  $A$ , viz obr. 12.13.

Pojmy vršek a stín poskytují intuitivní vztah mezi šedotónovou a binární morfologií. *Věta o homeomorfismu stínů* říká, že stín je homeomorfismem ze šedotónové do binární morfologie. Nechť  $F, K \subseteq \mathcal{E}^{n-1}$ ,  $f: F \rightarrow \mathcal{E}$  a  $k: K \rightarrow \mathcal{E}$ . Potom platí (důkaz lze najít např. v [26])

$$\begin{aligned} (a) \quad & U[f \oplus k] = U[f] \oplus U[k], \\ (b) \quad & U[f \ominus k] = U[f] \ominus U[k]. \end{aligned} \tag{12.34}$$

Homeomorfismus stínů lze použít pro odvození vlastností operací šedotónové morfologie. Operace se vyjádří pomocí stínů a vršků, převede se díky teorému o homeomorfismu stínů na operace binárními obrazy a nakonec se s využitím definic šedotónové eroze a dilatace převede



**Obrázek 12.19** Transformace vrchní část klobouku dovoluje najít objekty (zde jasnější než pozadí) na postupně se měnícím pozadí.

zpět na šedotónové obrazy. Tímto postupem lze ověřit, že v platnosti zůstávají vlastnosti, které už známe z binární morfologie jako komutativita dilatace, pravidlo řetězení dovolující rozložit složité strukturní elementy na postupné oprace s menšími strukturními elementy, dualita mezi erozí a dilatací, atd.

Šedotónové otevření a uzavření lze zavést stejným způsobem jako v binární morfologii. Šedotónové otevření je definováno jako  $f \circ k = (f \ominus k) \oplus k$ . Podobně pro šedotónové uzavření platí  $f \bullet k = (f \oplus k) \ominus k$ . Připomeňme, že označení  $\check{k}$  znamená transpozici množiny, tj. symetrickou množinu vzhledem k počátku souřadnic. Dualita mezi otevřením a uzavřením je vyjádřena jako

$$-(f \circ k)(x) = [(-f) \bullet \check{k}](x). \quad (12.35)$$

Šedotónové otevření obrazu  $f$  strukturním elementem  $k$  si lze názorně představit [26] jako posouvání funkce  $k$  krajinou  $f$ . Poloha všech nejvyšších bodů nějakou částí  $k$  při posunu dává otevření. Podobná interpretace existuje i pro erozi.

Šedotónové otevření a uzavření se používá k extrakci částí obrazu s daným tvarem a šedotónovou strukturou.

### 12.3.3 Transformace vrchní část klobouku

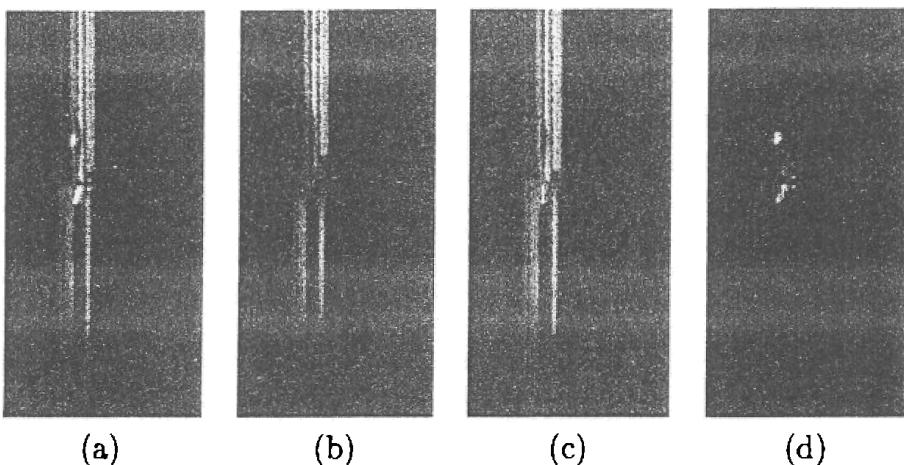
Mějme šedotónový obraz  $X$  a strukturní element  $K$ . Množinový rozdíl mezi otevřením a původním obrazem, tj.  $X \setminus (X \circ K)$  vytváří novou užitečnou operaci zvanou *vrchní část klobouku* (angl. top hat transformation) [64].

Transformace vrchní část klobouku se používá jako jednoduchý nástroj pro segmentaci objektů, které se v šedotónovém obrazu liší od pozadí, i když se jas pozadí pomalu mění. Části v obraze, které se nevejdou do strukturního elementu  $K$  použitého pro otevření, se odstraní. Po odečtení otevřeného obrazu od původního se ukáží právě jen odstraněné části obrazu. Ty lze v obraze najít jednoduchým prahováním. Myšlenky jsou ilustrovány na 1D průběhu na obr. 12.19. Zde si je možné domyslet i původ názvu. Transformace by našla pouze vrchní část klobouku, pokud by strukturní element byl větší než otvor pro hlavu.

Ještě lepší vlastnosti pro složitější průběh jasu na pozadí než transformace vrchní část klobouku má segmentace pomocí rozvodí (angl. watershed), kterou popíšeme v části 12.6.3.

Ukažme si příklad použití transformace vrchní část klobouku v průmyslu. Při výrobě teploměrů měřících maximální teplotu (např. běžný lékařský teploměr) je potřebné zúžit skleněnou kapiláru, aby se rtuť po ochlazení nemohla snadno vrátit do jímky. Kapilára se zužuje plynovým plamenem a je v ní přitom mírný podtlak. Při ohřátí se sklo zborgí směrem dovnitř. Když kapiláru zužoval člověk bez automatizace, bylo zužované místo vhodně osvětleno a odlesk

světla od skla se promítal na malé promítací plátno. Při borcení skla byl patrný odlesk, který zacvičené obsluze indikoval, že má mezi plynový hořák a kapiláru vložit plechovou clonu. Při automatizaci této technologické operace bylo potřebné odlesk detektovat z digitálního obrázku, na což byly použity operace matematické morfologie, viz obr. 12.20.

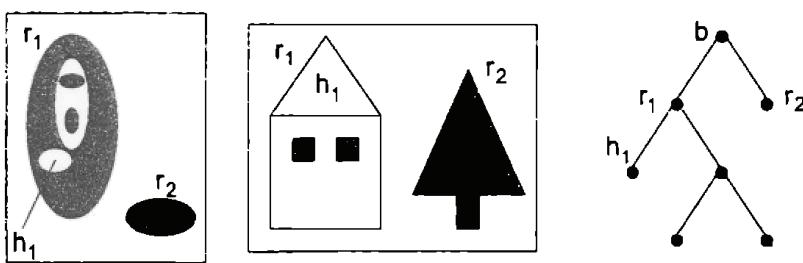


**Obrázek 12.20** Příklad na šedotónové otevření a transformaci vrchní část klobouku použité při automatizaci zužování kapilár skleněných teploměrů: (a) Výchozí obraz kapiláry  $512 \times 256$  pixelů. (b) Eroze strukturním elementem svislého tvaru o šířce jeden pixel a délce 20 pixelů. (c) Dilatace (tj. otevření výchozího elementu) týmž strukturním elementem. (d) Výsledná segmentace míst, kde došlo k odrazu světla při borcení kapiláry pomocí transformace vrchní část klobouku. Poděkování V. Smutný, R. Sára, FEL ČVUT, Praha a P. Kodl, Výzkumné středisko Rockwell Automation Praha.

## 12.4 Skelet a označování objektů

### 12.4.1 Transformace zachovávající homotopický strom

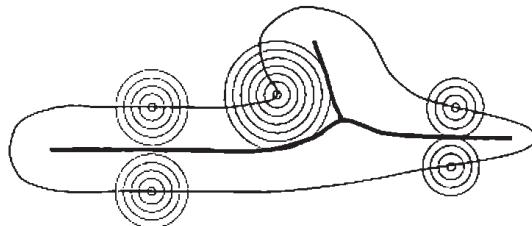
Topologické vlastnosti objektů se opírají o vlastnost souvislosti, jak už víme z kapitoly 7.3. Pro studium topologických vlastností se hodí i matematická morfologie, a to zejména homotopické transformace [87].



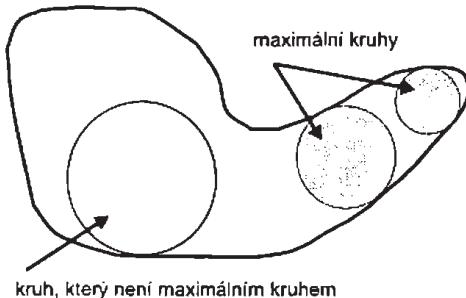
**Obrázek 12.21** Dva různé obrazy se stejným topologickým stromem.

Transformace je homotopická, když zachovává relaci souvislosti mezi oblastmi a děrami v obraze. Relaci souvislosti lze znázornit homotopickým stromem. Kořen stromu odpovídá pozadí. První větvení ukazuje na oblasti uvnitř pozadí. Druhé větvení odpovídá děram v objektech, atd.

Na obr. 12.21 je homotopický strom, který odpovídá dvěma různým obrazům. Obraz vlevo může představovat buňky v biologickém preparátu a obraz uprostřed domeček se smrkem. Oba obrazy mají týž homotopický strom. Kořen stromu  $b$  reprezentuje pozadí, uzel stromu  $r_1$



Obrázek 12.22 Skelet jako místa na suchém trávníku, kde se setkají dva a více šířící se ohně.



Obrázek 12.23 Vepsaný kruh a maximální vepsaný kruh v euklidovské rovině.

odpovídá větší svisle orientované buňce (obrysu domu), uzel  $r_2$  ukazuje na menší vodorovně orientovanou buňku (na smrk). Uzel  $h_1$  patří k prázdné díře v buňce  $r_1$  (díře ve střeše domečku). Význam dalších uzlů stromu by už měl být zřejmý.

Transformace je homotopická, když nemění homotopický strom. Třídu homotopických transformací si můžeme představit na příkladě. Představme si, že jsme něco namalovali lihovým fixem na pouťový nafukovací balónek. Nyní můžeme balónek všelijak natahovat, nakreslené objekty se budou různě deformovat, ale relace souvislosti a homotopický strom zůstanou zachovány. Ve zpracování obrazu mohou někdy studované objekty měnit tvar a přitom zachovají homotopický strom. Příkladem jsou buněčné struktury v mikroskopických obrazech.

#### 12.4.2 Skelet, maximální kruh

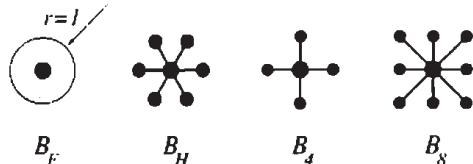
Někdy je výhodné vyjádřit podlouhlé objekty v obrazu pomocí "tenkých tyček", které tvoří skelet (kostru) objektu. Výklad pro názornost začneme ve spojitém euklidovském prostoru, což je názornější než v diskrétním rastru, na který dojde později.

S myšlenkou skeletu přišel ve zpracování obrazu Blum [3], tehdy pod názvem střední osa (angl. medial axis transform). Ideu znázornil následující představou o požáru trávníku. Představme si oblast (bodovou množinu)  $X \subset \mathbb{R}^2$ . Řekněme, že oblastí je suchý trávník a zapálíme jej v jednom okamžiku po celém obvodu. Předpokládejme, že se oheň se šíří konstantní rychlostí. Skelet množiny  $S(X)$  budou tvořit všechny body, kde se setkají dva a více ohnů, viz obr. 12.22.

Formální definice skeletu se opírá o pojem maximálního kruhu (ve 3D by to byla koule). *Maximální kruh*  $B(p, r)$  se středem  $p$  a poloměrem  $r$ ,  $r \geq 0$ , je množina bodů, jejichž vzdálenost  $d$  od středu je menší nebo rovna  $r$ .

Kruh  $B$  vepsaný do množiny  $X$  se nazývá *maximálním kruhem*, právě když se hranice množiny dotýká ve dvou a více bodech, což znamená, že pro dané místo dotyku již kruh nelze zvětšit. Představa vepsaného kruhu a maximálního vepsaného kruhu je znázorněna na obr. 12.23.

Pro diskrétní obrazy závisí definice vzdálenosti na použité mřížce a definici souvislosti.



**Obrázek 12.24** Kruhy o poloměru jedna pro různě definovanou vzdálenost v digitálním obrazu. Zleva euklidovská vzdálenost, 6-, 4-, a 8-souvislost.

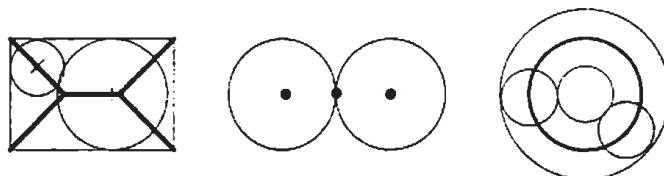
Různé kruhy o jednotkovém poloměru jsou znázorněny na obr. 12.24.

Rovina  $\mathcal{R}^2$  s euklidovskou vzdáleností implikuje kruh  $B_E$ . Pro diskrétní rovinu  $\mathcal{Z}^2$  se obvykle používají tři definice vzdálenosti. Pro hexagonální mřížku 6-souvislost získáme hexagonální kruh  $B_H$ . Je-li definičním oborem čtvercová mřížka, potom můžeme dostat dva kruhy  $B_4$  pro 4-souvislost a  $B_8$  pro 8-souvislost.

*Skelet pomocí maximálních kruhů  $S(X)$*  množiny  $X \subset \mathcal{Z}^2$  je innožina středů  $p$  maximálních kruhů:

$$S(X) = \{p \in X : \exists r \geq 0, B(p, r) \text{ je maximální kruh množiny } X\}.$$

Tato definice je v euklidovské rovině přirozená. Skeletem kruhu je jeho střed, proužek se zaoblenými konci je převeden na tenkou čáru uprostřed proužku, atd.



**Obrázek 12.25** Skelet obdélníka, dvou dotýkajících se kruhů, a mezikruží.

Na obr. 12.25 je ukázáno několik objektů se skelety. Všimněte si, že euklidovský skelet dvou dotýkajících se kruhů je tvořen třemi body, dva jsou ve středu kruhů a třetí v místě jejich dotyku, protože sem lze vepsat nekončně malý kruh, který se dotýká obou kruhů. Možná, že bychom od kostry intuitivně očekávali, že zde bude tvořena čarou spojující oba středy kruhů.

Skelet zavedený pomocí vpisovaných maximálních kruhů má z hlediska aplikací dvě nepříznivé vlastnosti. Za prvé, ne vždy zachovává homotopii původní množiny (viz skelet dotýkajících se kruhů). Za druhé, v diskrétní mřížce mohou vycházet skelety s tloušťkou větší než jeden pixel. Později uvidíme, že lze skelet definovat pomocí sekvenčního zeslabování, a tak se vyhnout dvěma zmíněným nepříznivým vlastnostem.

Dilataci můžeme opřít o jednu ze tří zmíněných definic souvislosti, a tak získat kruhy o různém poloměru. Nechť  $nB$  označuje kruh o poloměru  $n$ :

$$nB = B \oplus B \oplus \dots \oplus B. \quad (12.36)$$

Skelet pomocí maximálních kruhů lze stanovit jako sjednocení rezidu transformace otevření množiny  $X$  pro různá měřítka (zde dána poloměrem kruhu) [87]

$$S(X) = \bigcup_{n=0}^{\infty} ((X \ominus nB) \setminus (X \ominus nB) \circ B). \quad (12.37)$$

Problémem je, že je výsledný skelet nesouvislý, a tudíž v mnoha aplikacích nepoužitelný. Proto se většinou dává přednost *homotopickému skeletu*, který z definice souvislost zachovává. V příšt-

tím části 12.4.3 ukážeme, jak taková homotopická skeletonizace založená na opakovaném ztenčování pracuje.

### 12.4.3 Ztenčování, ztlušťování a homotopický skelet

Jednou aplikací operace tref či mříž popsané v části 12.2.3 je *ztenčování* a *ztlušťování* bodo-vých množin. Nechť  $X$  je obraz a  $B = (B_1, B_2)$  je strukturní element. Upozorněme, že  $B$  zde neoznačuje kruh. *Ztenčování* je definováno jako

$$X \oslash B = X \setminus (X \otimes B). \quad (12.38)$$

*ztlušťování* se definuje

$$X \odot B = X \cup (X \otimes B). \quad (12.39)$$

Při ztenčování se část objektu určená strukturním elementem  $B$  množinově odečítá od objektu samého. Při ztlušťování se objekt sjednocuje s částí pozadí danou strukturním elementem  $B$ . Ztenčování a ztlušťování jsou duální transformace

$$(X \odot B)^c = X^c \oslash B, \quad B = (B_2, B_1). \quad (12.40)$$

Ztenčování a ztlušťování se velmi často používají opakovaně. Nechť  $\{B_{(1)}, B_{(2)}, B_{(3)}, \dots, B_{(n)}\}$  označuje posloupnost složených strukturních elementů  $B_{(i)} = (B_{i1}, B_{i2})$ . *Sekvenční ztenčování* může být pro čtvercový rastr vyjádřeno pomocí posloupnosti 8 strukturních elementů

$$X \oslash \{B_{(i)}\} = (((X \oslash B_{(1)}) \oslash B_{(2)}) \dots \oslash B_{(n)}). \quad (12.41)$$

a *sekvenční ztlušťování* jako

$$X \odot \{B_{(i)}\} = (((X \odot B_{(1)}) \odot B_{(2)}) \dots \odot B_{(n)}). \quad (12.42)$$

Existuje několik posloupností strukturních elementů  $\{B_{(i)}\}$ , které jsou z praktického pohledu užitečné. Jejich vlastností je, že při ztenčování neporuší souvislost objektu. Uvedeme jen dvě z nich pro oktagonální rastr, a to posloupnosti z tzv. *Golayovy abecedy* [18]. Strukturní elementy rozměru  $3 \times 3$  uvedeme ve dvou základních polohách, ostatní polohy si čtenář domyslí pootočením v příslušném diskrétním rastru. Pro hexagonální rastr jsou elementy Golayovy abecedy uvedeny v [87].

Kvůli stručnosti zápisu vyjádříme složený strukturní element jedinou maticí. Hodnota 1 v matici znamená, že příslušný prvek patří  $B_1$  (v transformaci tref či mříž se tedy ověřuje příslušnost k objektu), hodnota 0 říká, že prvek patří do  $B_2$  (požaduje se příslušnost k pozadí) a konečně hodnota \* znamená, že prvek se ve srovnávání neúčastní a jeho hodnota nehraje roli.

Operace ztenčování a ztlušťování jsou idempotentní, tj. po určitém počtu kroků se přestanou měnit. Použijí se iteračně a skutečnost, že obrazy ve dvou po sobě následujících iteracích jsou totožné, indikuje, že se iterace mají ukončit.

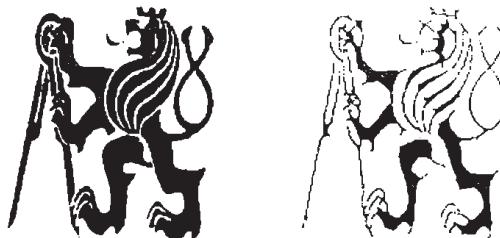
#### Sekvenční ztenčování strukturním elementem $L$ Golayovy abecedy

Toto sekvenční ztenčování je pro jednoduchost používáno pro hledání homotopické háhrady skeletu. Výsledný skelet se skládá pouze z čar tloušťky 1 a izolovaných bodů.

Strukturní element  $L$  z Golayovy abecedy je

$$L_1 = \begin{bmatrix} 0 & 0 & 0 \\ * & 1 & * \\ 1 & 1 & 1 \end{bmatrix}, \quad L_2 = \begin{bmatrix} * & 0 & 0 \\ 1 & 1 & 0 \\ * & 1 & * \end{bmatrix} \dots \quad (12.43)$$

(Ostatních šest poloh si čtenář pootočí sám). Obr. 12.26 ukazuje výsledek ztenčování strukturním elementem  $L$  po pěti iteracích, aby byl patrný mezivýsledek. Obr. 12.27 výsledek ztenčení po dosažení idempotence.



Obrázek 12.26 Sekvenční ztenčování pomocí elementu  $L$  po 5 iteracích. Výchozí obraz je vlevo.



Obrázek 12.27 Homotopická náhrada skeletu (element  $L$ ). Výchozí obraz je vlevo.

### Sekvenční ztenčování strukturním elementem $E$ Golayovy abecedy

Předpokládejme, že jsme již vypočítali homotopickou náhradu skeletu pomocí elementu  $L$ . Nalezený skelet má díky ostrým rohům ve výchozích objektech roztržené konce, vzpomněte si, jak vypadá skelet pro obdélník. Je-li ztenčování použito na tenké objekty, např. tenké čáry v technickém výkresu, potom by mohlo být žádoucí tyto roztržené konce odstranit. K tomu slouží sekvenční ztenčování elementem  $E$ , které zkracuje volné konce skeletu. S uzavřenými liniemi nedělá nic. Pokud by se ztenčování nechalo běžet až do dosažení idempotence, zůstaly by v obraze pouze uzavřené linie.

Strukturní element  $E$  z Golayovy abecedy je opět dán osmi maskami rozměru  $3 \times 3$  pro osm možných pootočení. My zde uvádíme jen první dvě

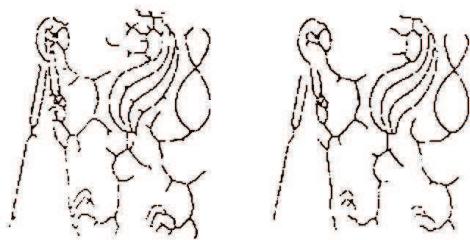
$$E_1 = \begin{bmatrix} * & 1 & * \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & * & * \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (12.44)$$

Obr. 12.28 ukazuje sekvenční ztenčování elementem  $E$  po 5 iteracích. Výchozím obrazem byl skeleton z obr. 12.27 vpravo. Všimněte si zkrácení volných konců.

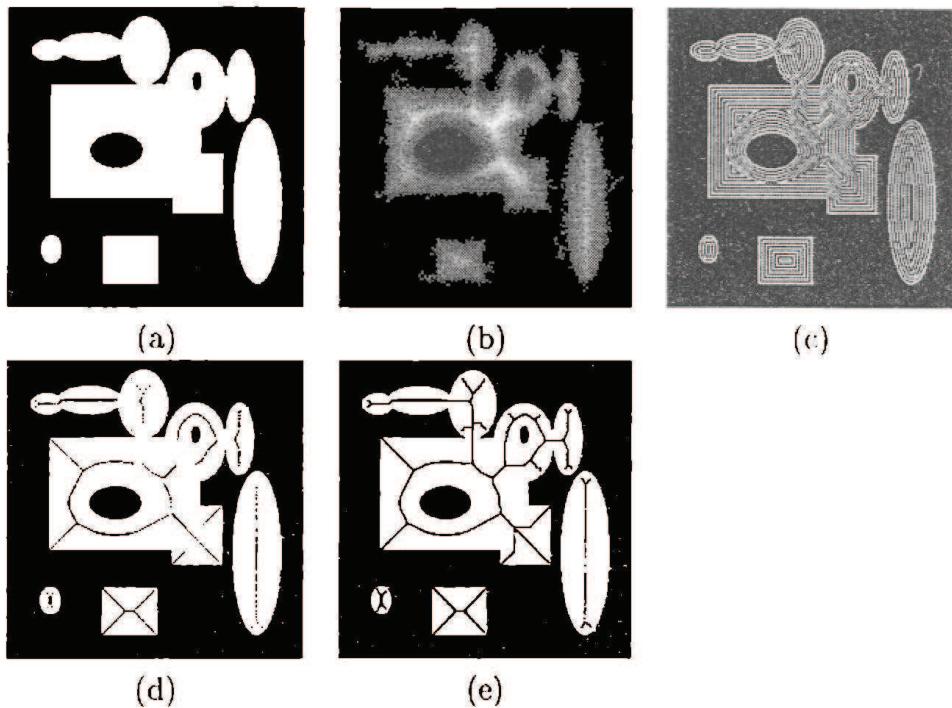
V Golayově abecedě [18] jsou ještě další elementy  $M$ ,  $D$ ,  $C$ . Dnes se již nepoužívají, protože existují lepší morfologické algoritmy pro výpočet skeletu, konvexního obalu a homotopických značek.

Výpočetně nejfektivnější algoritmus pro výpočet homotopického skeletu pracující s obrazem v nekomprimovaném tvaru, o kterém víme, je Vincentův algoritmus na základě maximálních kruhů [93]. Jeho výsledky jsou na obr. 12.29.

Ukažme ještě na příkladě, že skelet lze také počítat pro trojrozměrné objekty. Obr. 12.30 ukazuje ztenčování 3D bodové množiny.



**Obrázek 12.28** Ztenčování elementem  $E$ . Vlevo je výchozí skelet z obr. 12.27. Vpravo je výsledek odstraňování volných konců po 5 iteracích.



**Obrázek 12.29** Výsledek rychlého algoritmu (L. Vincent) pro výpočet skeletu pomocí maximálních kruhů: (a) výchozí binární obraz, (b) vzdálenostní funkce (bude vysvětlena dál), (c) vzdálenostní funkce znázorněná vrstevnicemi, (d) nesouvislý skelet pomocí maximálních kruhů, (e) výsledný skelet.

#### 12.4.4 Značkování oblastí, konečná eroze

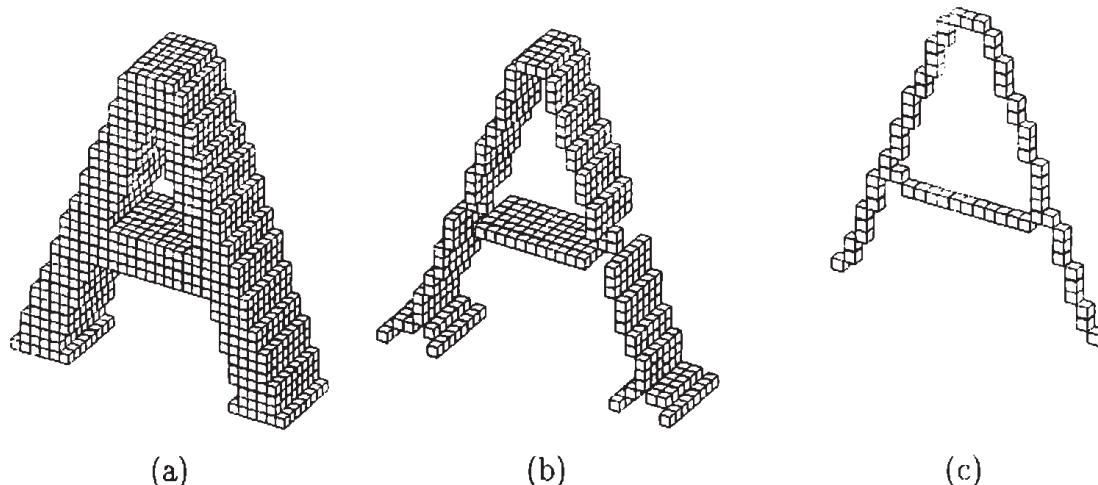
V binární morfológii lze bodovou množinu  $X$  také vyjádřit jako sjednocení maximálních kruhů  $B$ . Každému bodu  $p$  skeletu  $S(X)$  je jednoznačně přiřazen maximální kruh o poloměru  $q_X(p)$  (angl. quench function<sup>1</sup>). Známe-li pro každý bod skeletu  $q_X(p)$ , což při skeletonizaci můžeme zjistit a zapamatovat, potom lze výchozí bodovou množinu  $X$  rekonstruovat jako sjednocení maximálních kruhů  $B$

$$X = \bigcup_{p \in S(X)} (p + q_X(p)B). \quad (12.45)$$

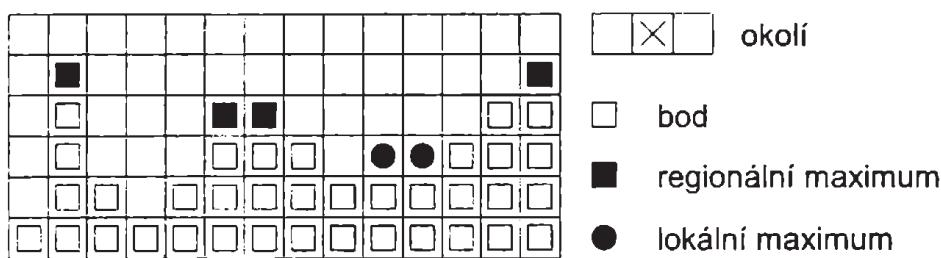
Poznamenejme, že předchozí vztah také dovoluje bezetrátovou kompresi binárních obrazů. Tyto myšlenky jsou použity pro kódování dokumentů v kompresním standardu "group 4" mezinárodní telekomunikační asociace CCITT.

Nyní bude výhodné, když rozlišíme několik druhů extrémů (šedotónové obrazové) funkce v daném definičním oboru. Pro zvýšení názornosti se budeme při výkladu opět opírat o to-

<sup>1</sup>V.H. Nenašel jsem výstižný český překlad a nechci zavést matoucí.



**Obrázek 12.30** Morfologické ztenčování ve 3D: (a) výchozí 3D bodová množina, písmeno A, (b) ztenčování v jednom směru, (c) výsledný skelet o tloušťce jeden voxel získaný ztenčováním obrazu (b) ve druhém směru. Poděkování K. Palágyi, University of Szeged, Maďarsko.



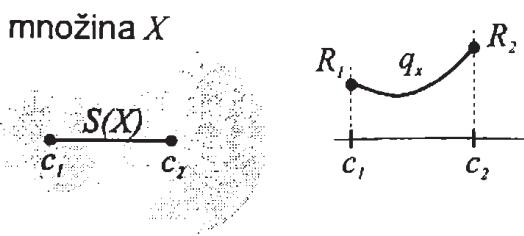
**Obrázek 12.31** 1D znázornění regionálního a lokálního maxima.

pografický pohled. *Globálním maximumm* je pixel s nejvyšší hodnotou, tj. nejsvětlejší bod nebo nejvyšší vrchol v krajině. Podobně *globální minimum* odpovídá nejtmařímu pixelu nebo nejhlbší prohlubni v krajině.

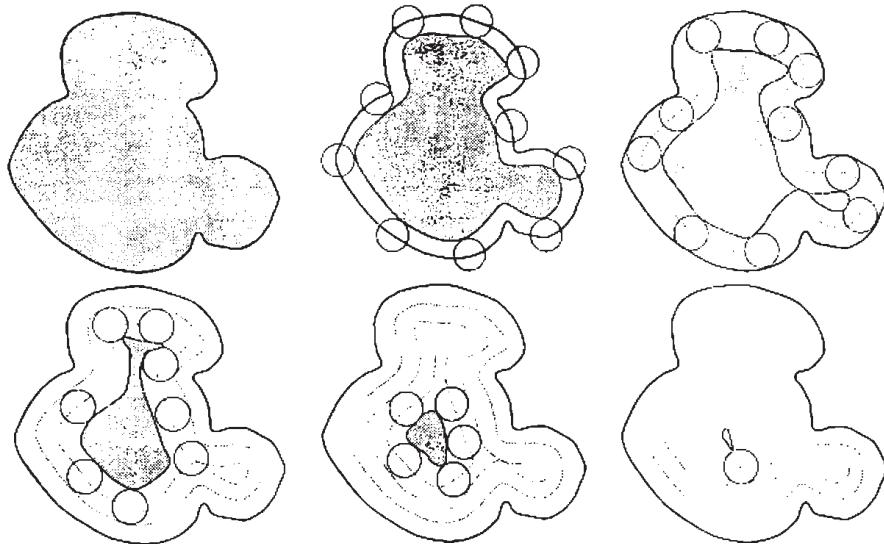
Pixel  $p$  v šedotónovém obrazu je *lokálním maximumm*, právě když pro každý sousední pixel  $q$  pixelu  $p$  platí  $I(p) \geq I(q)$ . V krajině pojem lokální může znamenat, že extrém lze studovat jen v malém okolí bodu, kde právě stojíme, např. pro hustou mlhu. V matematické morfologii je malé okolí definováno strukturním elementem. Pokud v okolí zkoumaného bodu není vidět pokles (sestup v krajině), jedná se o lokální maximum. Lokální minimum bychom zavedli analogicky.

*Regionální maximum M* vzorkované a kvantované funkce (digitálního šedotónového obrazu)  $I$  je souvislá množina pixelů s odpovídající hodnotou  $h$  (platí ve výšce  $h$ ), kde každý pixel souřadící s množinou  $M$  má menší hodnotu než  $h$ . Z regionálního maxima neexistuje cesta vzhůru. Z topografického hlediska odpovídají regionální extrémy vrcholům a prohlubním v krajině. Každý pixel regionálního maxima  $M$  v obrazové funkci  $I$  je také lokálním maximem. Opak neplatí, tj. existují lokální maxima, která nejsou regionálními maximy. Lokální a regionální maxima jsou ilustrována v 1D případě na obr. 12.31.

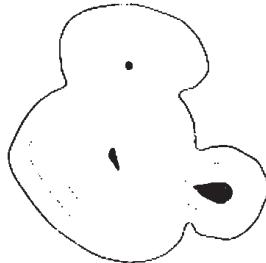
Nyní si představme, že bychom chtěli každou konvexní oblast v binárním obrazu reprezentovat jedinou značkou, která leží v nějakém smyslu uprostřed oblasti. Pokud by oblastmi byly nedotýkající se kruhy, byl by úkol triviálně splněn. Značkami by byly středy maximálních kruhů. Obecná situace není tak jednoduchá, protože konvexní oblast je složitější než kruh. Pro



**Obrázek 12.32** Skelet bodové množiny a odpovídající funkce  $q_X(p)$  přiřazující pixelu  $p$  poloměr maximálního kruhu  $q_X(p)$ . Regionální maxima  $R_1$ ,  $R_2$  funkce  $q_X(p)$  odpovídají výsledku konečné eroze.



**Obrázek 12.33** Při opakované erozi se množina  $X$  nejprve rozdělí na oddělené konvexní oblasti, které postupně zmizí. Pokud si zapamatujeme rezidua každé oblasti, tj. poslední bodovou množinu před jejím zmizením, tvoří jejich sjednocení výsledek konečné eroze  $Ult(X)$ .



**Obrázek 12.34** Konečná eroze je tvořena rezidui oblastí, tj. oblastmi těsně před zmizením při opakovaných erozích.

konvexní oblast vyjděme z představy, že objekt budeme postupně erodovat strukturálním elementem odpovídajícím kruhu o jednotkovém poloměru. Značkou oblasti bude poslední zbylá oblast před tím, než zcela zmizí. Budeme-li opakovaně erodovat nekonvexní oblast, potom se nejprve přirozeně rozdělí na jednotlivé konvexní části a dále platí předchozí úvaha.

Cestou k formalizaci předchozí představy vytváření značek je analýza funkce  $q_X(p)$ , která každému bodu skeletu  $p$  přiřazuje poloměr maximálního kruhu. Právě kvůli její analýze jsme rozlišili mezi regionálními a lokálními maximy. Analýza funkce  $q_X(p)$  se nám bude hodit pro zavedení konečné eroze, která poskytne značky objektů v binárních obrazech.

*Konečná eroze* bodové množiny  $X$  se označuje  $Ult(X)$  a je definována jako množina skládající se z regionálních maxim funkce  $q_X(p)$ . Uvažujme nejprve případ jednoduché nekonvexní množiny  $X$ , která se skládá ze dvou překrývajících se kruhů, viz obr. 12.32. Skelet odpovídá úsečce spojující středy kruhů. Funkce  $q_X(p)$  má dvě regionální maxima ležící ve středu výchozích kruhů. Tato maxima odpovídají konečné erozi, tj. značkám. Konečná eroze poskytne značky i pro obecné nekonvexní objekty.

Jako ilustrační příklad uvažujme nekonvexní oblast  $X$ , tj. bodovou množinu v binárním obraze, která sestává ze tří konvexních překrývajících se konvexních složek různé velikosti. Při opakované erozi strukturním elementem ve tvaru jednotkového kruhu se oblast nejprve změní, potom rozdělí na složky a nakonec zmizí, viz obr. 12.33. Sjednocení dílčích oblastí těsně před zmizením poskytne konečnou erozi, tj. hledané značky, viz obr. 12.34.

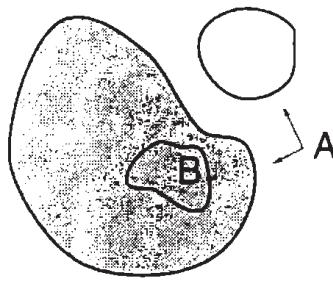
#### 12.4.5 Vzdálenostní funkce a efektivní výpočet konečné eroze

Nyní se zaměříme na efektivní výpočet konečné eroze a upozorníme na souvislost s postupem morfologické rekonstrukce. Předpokládejme dvě bodové množiny  $A, B, B \subseteq A$ . Morfologická rekonstrukce  $\rho_A(B)$  množiny  $A$  z množiny  $B$  je sjednocením souvislých částí množiny  $A$  s neprázdným průnikem s  $B$  (viz obr. 12.35, všimněte si, že se množina  $A$  skládá ze dvou souvislých částí). Množina  $B$  se obvykle skládá ze značek, které dovolí rekonstrukci původní množiny. Morfologickou rekonstrukci podrobň rozebereme v části 12.4.7.

Když jsme zavedli rekonstrukci  $\rho$ , můžeme vyjádřit vzorec pro výpočet konečné eroze. Nechť  $\mathcal{N}$  je množina přirozených čísel, která poslouží pro rostoucí poloměry kruhů. Konečná eroze se může vyjádřit jako

$$Ult(X) = \bigcup_{n \in \mathcal{N}} ((X \ominus nB) \setminus \rho_{X \ominus nB}(X \ominus (n+1)B)) . \quad (12.46)$$

Efektivní výpočet konečné eroze se opírá o vzdálenostní funkci (která je ostatně jádrem několika rychlých morfologických algoritmů, jak dále uvidíme). Vzdálenostní funkce  $dist_X(p)$  přiřazuje



**Obrázek 12.35** Rekonstrukce  $\rho_A(B)$  (znázorněno šedivě) množiny  $A$  z množiny známkového jádra  $B$ . Všimněme si, že se množina  $A$  skládá z více než jedné souvislé oblasti.

každému pixelu  $p$  z množiny  $X$  velikost první eroze množiny, která už neobsahuje pixel  $p$ , tj.

$$\forall p \in X, \quad dist_X(p) = \min\{n \in \mathcal{N}, p \text{ not in } (X \ominus nB)\}. \quad (12.47)$$

Vzdálenostní funkce se chová podle očekávání, tj.  $dist_X(p)$  je nejkratší vzdálenost mezi pixelem  $p$  a doplňkem množiny  $X^C$ .

Vzdálenostní funkce má dvě přímá použití:

- Konečná eroze množiny  $X$  je tvořena sjednocením regionálních maximálních vzdálenostních funkcí množiny  $X$ .
- Skelet pomocí maximálních kruhů množiny  $X$  je dán množinou lokálních maximálních vzdálenostních funkcí množiny  $X$ .

Posledním pojmem zavedeným v této části je *skelet pomocí zón vlivu* (angl. skeleton by influence zones, zkracováno SKIZ). Nechť je množina  $X$  složena z  $n$  souvislých částí  $X_i$ ,  $i = 1, \dots, n$ . Zóna vlivu  $Z(X_i)$  obsahuje body, ležící k množině  $X_i$  blíže než k dalším souvislým částem  $X$ .

$$Z(X_i) = \{p \in \mathbb{Z}^2, \forall i \neq j, d(p, X_i) \leq d(p, X_j)\}. \quad (12.48)$$

SKIZ - skelet pomocí zón vlivu označený  $SKIZ(X)$  je dán množinou hranic zón vlivu  $\{Z(X_i)\}$ .

#### 12.4.6 Geodetické transformace

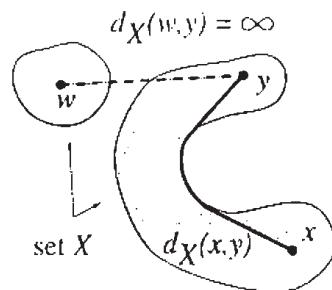
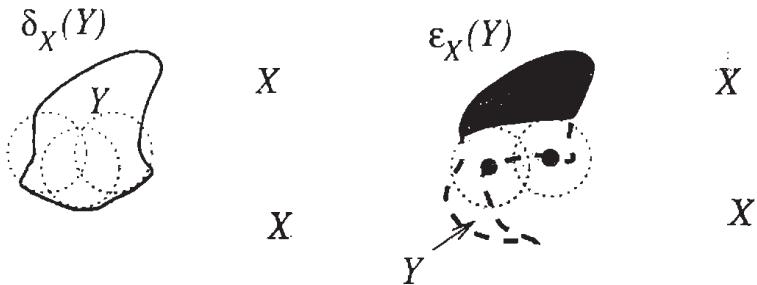
Geodetické metody [96] změní morfologické operace tak, aby operovaly pouze na části obrázku. Má-li se například rekonstruovat objekt ze značky, řekněme buněčného jádra, je žádoucí zabránit růstu mimo buňku. Další důležitou výhodou je, že se strukturní element může měnit v každém pixelu podle lokálních hodnot obrazové funkce.

Ústředním pojmem geodetických metod je geodetická vzdálenost. Cesta mezi dvěma body je omezena a má ležet uvnitř nějaké množiny. Tento přístup je vlastní staré disciplíně, geodézii, která měří vzdálenosti na povrchu Země. Představme si, že chceme stanovit nejkratší cestu mezi Prahou a Sydney. Nejkratší vzdálenost míří nitrem Země. Cestovatel se zajímá o geodetickou vzdálenost, tj. takovou, která je na povrchu nebo blízko povrchu Země.

Geodetická vzdálenost  $d_X(x, y)$  je délka nejkratší cesty mezi dvěma body  $x, y$ , za podmínky, že leží uvnitř množiny  $X$ . Není-li taková cesta, definuje se geodetická vzdálenost  $d_X(x, y) = +\infty$ . Geodetická vzdálenost je ilustrována na obr. 12.36.

Geodetický kruh (koule, nadkoule pro vícedimenzionální prostor) je kruh omezený množinou  $X$ . Geodetický kruh  $B_X(p, n)$  se středem  $p \in X$  a poloměrem  $n$  je definován jako

$$B_X(p, n) = \{p' \in X, d_X(p, p') \leq n\}. \quad (12.49)$$

Obrázek 12.36 Geodetická vzdálenost  $d_X(x, y)$ .Obrázek 12.37 Znázornění geodetické dilatace (vlevo) a eroze (vpravo) množiny  $Y$  obsažené v množině  $X$ .

Zavedení geodetického kruhu dovoluje použít eroze a dilatace pouze uvnitř podmnožiny  $Y$  obrazu  $X$ .

*Geodetická dilatace*  $\delta_X^{(n)}$  o velikosti  $n$  množiny  $Y$  obsažené v množině  $X$  je definována

$$\delta_X^{(n)}(Y) = \bigcup_{p \in Y} B_X(p, n) = \{p' \in X, \exists p \in Y, d_X(p, p') \leq n\}. \quad (12.50)$$

Podobně je zavedena duální transformace *geodetická eroze*  $\epsilon_X^{(n)}(Y)$  velikosti  $n$  podmnožiny  $Y$  obsažené v množině  $X$

$$\epsilon_X^{(n)}(Y) = \{p \in Y, B_X(p, n) \subseteq Y\} = \{p \in Y, \forall p' \in X \setminus Y, d_X(p, p') > n\}. \quad (12.51)$$

Geodetická dilatace a eroze je ilustrována na obr. 12.37.

Výsledek geodetické operace na množině  $Y \subseteq X$  je vždy obsažen v  $X$ . Z hlediska implementace je rozumné napsat nejednodušší geodetickou dilataci velikosti 1, tj.  $\delta_X^{(1)}$  množiny  $Y \subset X$  jako průnik dilatací množiny  $Y$  (pomocí jednotkového kruhu  $B$ ) s množinou  $X$ .

$$\delta_X^{(1)} = (Y \oplus B) \cap X. \quad (12.52)$$

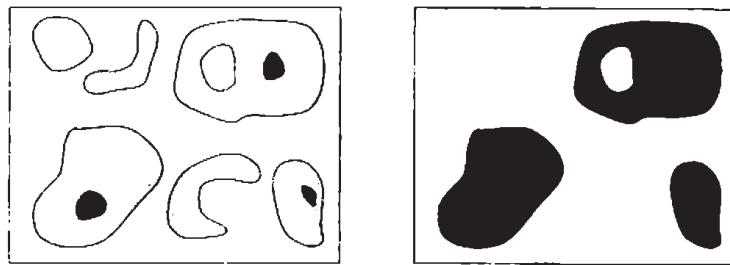
Geodetické dilatace větší velikosti se získají postupným skládáním  $n$ -krát použité dilatace velikosti 1

$$\delta_X^{(n)} = \underbrace{\delta_X^{(1)}(\delta_X^{(1)}(\delta_X^{(1)} \dots (\delta_X^{(1)}))))}_{n\text{-krát}}. \quad (12.53)$$

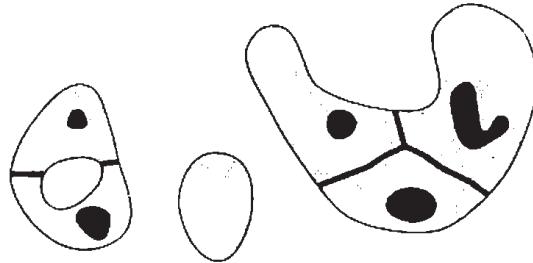
Pro geodetickou erozi je rychlá implementace analogická.

### 12.4.7 Morfologická rekonstrukce

Předpokládejme, že chceme rekonstruovat objekty daného tvaru z binárního obrazu, který byl získán prahováním z šedotónového obrazu. Sjednocení všech souvislých oblastí tvoří množinu



**Obrázek 12.38** V levé části je výchozí obraz pro rekonstrukci množiny  $X$  (znázorněno šedivě) ze značek  $Y$  (znázorněny černě). V pravé části je výsledek rekonstrukce (černě).



**Obrázek 12.39** Geodetické zóny vlivu.

$X$ . Ale jen některé z těchto oblastí byly označeny značkami, které tvoří množinu  $Y$ . Úlohou je rekonstruovat jen označené objekty, jak je ilustrováno na obr. 12.38.

Opakování geodetické dilatace množiny  $Y \subset X$  umožňuje rekonstrukci souvislých částí  $X$ , které byly původně označeny pomocí  $Y$ . Při dilataci se značka nespojí s jinou souvislou složkou  $X$ , ve které nebyla značka z  $Y$ . Takové souvislé části zmizí. Postup geodetického dilatování skončí, když je dosaženo idempotence.

$$\forall n > n_0, \delta_X^{(n)}(Y) = \delta_X^{(n_0)}(Y). \quad (12.54)$$

Tato operace se nazývá *morfologická rekonstrukce* a označuje se  $\rho_X(Y)$ ,

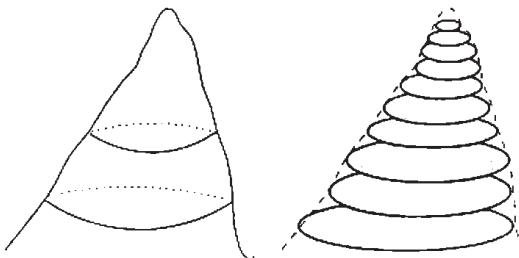
$$\rho_X(Y) = \lim_{n \rightarrow \infty} \delta_X^{(n)}(Y). \quad (12.55)$$

Někdy je v aplikacích žádoucí, aby mohla být jedna souvislá komponenta označena více značkami z  $Y$  a současně nebylo možné, aby při se při geodetické dilataci jednotlivé narůstající oblasti spojily. Pojem zóny vlivu z části 12.4.5 lze zobecnit na *geodetickou zónu vlivu* souvislé části  $Y$  uvnitř množiny  $X$ . Myšlenka je znázorněna na obr. 12.39. Růst ze značek se ukončí, když by se oblasti měly spojit.

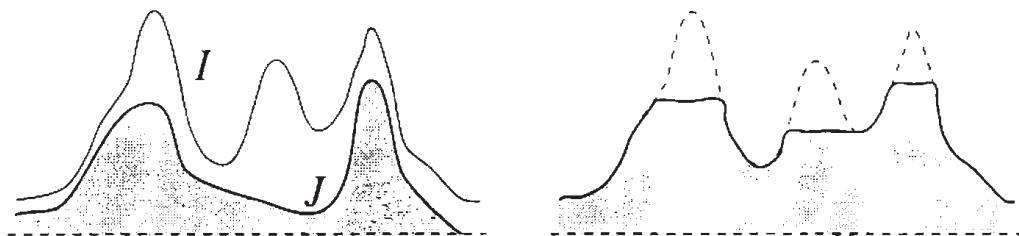
Nyní můžeme přistoupit k zobecnění morfologické rekonstrukce na šedotónové obrazy. Budeme potřebovat rozšířit geodetické pojmy na obrazovou funkci s více hodnotami jasu. Základem rozšíření je tvrzení platící pro diskrétní obrazy, a to, že každá rostoucí transformace definovaná pro binární obrazy je rozšiřitelná pro šedotónové obrazy [87]. Myslíme transformaci  $\Psi$ , pro niž platí

$$\forall X, Y \subset \mathcal{Z}^2, Y \subseteq X \implies \Psi(Y) \subseteq \Psi(X). \quad (12.56)$$

Transformaci  $\Psi$  lze zobecnit tak, že se na šedotónový obraz budeme dívat jako na na sebe položené binární obrazy získané prahováním s postupně rostoucím prahem. Této představě se říká dekompozice obrazu  $I$  pomocí prahování [58]. Nechť  $D_I$  je definiční obor obrazu  $I$  a nechť hodnoty jasu obrazu  $I$  jsou  $\{0, 1, \dots, N\}$ . Dekompozici obrazu pomocí prahování můžeme formálně



Obrázek 12.40 Dekompozice šedotónového obrazu pomocí prahování.



Obrázek 12.41 1D znázornění šedotónové rekonstrukce masky  $I$  ze značky  $J$ .

zapsat

$$T_k(I) = \{p \in D_I, \quad I(p) \geq k\}, \quad k = 0 \dots N. \quad (12.57)$$

Myšlenka dekompozice obrazu pomocí prahování je znázorněna na obr. 12.40. Obraz dekomponovaný pomocí prahování  $T_k(I)$  splňuje relaci inkluze

$$\forall k \in [1 \dots N], \quad T_k(I) \subseteq T_{k-1}(I). \quad (12.58)$$

Uvažujme rostoucí transformaci  $\Psi$  použitou na každý binární obraz získaný dekompozicí pomocí prahování. Relace inkluze je splněna. Transformace  $\Psi$  můžeme rozšířit na šedotónové obrazy s využitím následujícího *principu dekompozice prahováním*,

$$\forall p \in D_I, \quad \Psi(I)(p) = \max\{k \in [0, \dots, N], \quad p \in \Psi(T_k(I))\}. \quad (12.59)$$

Vraťme se k morfologické rekonstrukci. *Binární geodetická rekonstrukce*  $\rho$  je rostoucí transformací a splňuje

$$Y_1 \subseteq Y_2, \quad X_1 \subseteq X_2, \quad Y_1 \subseteq X_1, \quad Y_2 \subseteq X_2, \quad \Rightarrow \rho_{X_1}(Y_1) \subseteq \rho_{X_2}(Y_2). \quad (12.60)$$

Zobecněme binární rekonstrukci na *šedotónovou morfologickou rekonstrukci*, a to s využitím principu dekompozice pomocí prahování podle rovnice (12.59). Nechť  $J, I$  jsou dva šedotónové obrazy s definičním oborem  $D$ , hodnotami jasu v diskrétním intervalu  $[0, 1, \dots, N]$ . Platí-li pro každý pixel  $p \in D$ ,  $J(p) \leq I(p)$ , potom je šedotónová rekonstrukce  $\rho_I(J)$  obrazu  $I$  na základě obrazu  $J$  (odpovídá značkám) dána jako

$$\forall p \in D, \quad \rho_I(J)(p) = \max\{k \in [0, N], \quad p \in \rho_{T_k}(T_K(J))\}. \quad (12.61)$$

Připomeňme, že při binární morfologické rekonstrukci narůstají souvislé části v masce, která je označena alespoň jedním pixelem (značkou). Pro šedotónové obrazy mají stejný význam špičky signálu. Šedotónová morfologická rekonstrukce zachová ty špičky, které jsou označeny pomocí  $J$ , viz obr. 12.41.

## 12.5 Granulometrie

Granulometrii zavedli stereologové (matematici snažící se rekonstruovat 3D tvar z řezů). Jméno má kořeny v latině, ve které *granulum* znamená zrno. Matheron [60] granulometrii použil k analýze porézních materiálů. Distribuci velikosti pórů kvantifikoval posloupností otevření a uzavření se strukturním elementem o rostoucí velikosti. Dnes je granulometrie důležitým nástrojem matematické morfologie, zvláště při analýze materiálů a v biologii. Hlavní výhodou granulometrie je, že dovoluje vyvodit informaci o měřítku, aniž by byla potřebná interpretace obrazu. Z hlediska předmětu zpracování signálů a obrazů je zajímavé, že analýza granulometrického spektra je podobná analýze frekvenčního spektra.

Začneme analogií – *prosíváním*. Předpokládejme, že vstupem je hromada kamenů (nebo granulí) o různých velikostech. Úlohou je zjistit, kolik kamenů patří do jednotlivých tříd daných velikostí. Tato úloha se řeší postupným prosíváním hromady sítem s rostoucí velikostí děr. Výsledkem analýzy je diskrétní funkce. Nezávisle proměnnou je velikost děr v sítu a závisle proměnnou je počet kamenů příslušné velikosti. V matematické morfologii budeme této funkci říkat *granulometrické spektrum* nebo *granulometrická křivka*. V binární matematické morfologii se prosívání nahrazuje opakováním otevření strukturním elementem s rostoucí velikostí.

V morfologii hraje granulometrie podobně významnou roli jako frekvenční analýza ve zpracování signálu nebo obrazu. Vzpomeňme, že ve frekvenčním vyjádření je signál tvořen lineární kombinací harmonických signálů o rostoucí frekvenci. *Frekvenční spektrum* udává, jak přispívají jednotlivé harmonické signály. Je zřejmé, že granulometrické spektrum je analogií k frekvenčnímu spektru.

Nechť  $\Psi = (\psi_\lambda)$ ,  $\lambda \geq 0$  je třída transformací závislých na parametru  $\lambda$ . Třída je *granulometrií* právě, když platí následující vlastnost transformace  $\psi$ :

$$\forall \lambda \geq 0, \quad \psi_\lambda \text{ je rostoucí}, \quad (12.62)$$

$$\psi_\lambda \text{ je antiextenzivní}. \quad (12.63)$$

$$\forall \lambda \geq 0, \mu \geq 0, \quad \psi_\lambda \psi_\mu = \psi_\mu \psi_\lambda = \psi_{\max(\lambda, \mu)}. \quad (12.64)$$

Důsledkem vlastnosti (12.64) je fakt, že pro každé  $\lambda \geq 0$  je transformace  $\psi_\lambda$  idempotentní. Transformace  $(\psi_\lambda)$ ,  $\lambda \geq 0$  vytváří třídu rostoucích otevření (přesněji algebraických otevření [87]), které zde zavedené otevření zobecňují). Lze ukázat, že pro konvexní strukturní element  $B$  tvoří otevření vzhledem k  $\lambda B = \{\lambda b, b \in B\}$ ,  $\lambda \geq 0$  granulometrii [61].

Uvažujme intuitivnější granulometrii pracující na diskrétních binárních obrazech (bodových množinách). Zde je granulometrií posloupnost otevření  $\psi_n$  indexovaná přirozeným číslem  $n \geq 0$ , kde výsledek dalšího otevření v posloupnosti je menší než předchozí. Vzpomeňte na analogické prosívání kamenů, kde velikost síta odpovídá velikosti otevření. Každé otevření z obrazu odstraní více než předchozí, až nakonec nezbyde nic. Každý krok prosívání charakterizuje míra  $m(X)$  na množině (obrazu)  $X$  (tj. počet pixelů ve 2D obrazu, voxelů ve 3D). Množinu charakterizuje rychlosť prosívání. Granulometrické spektrum je takovou charakteristikou.

*Granulometrická funkce* zvaná také *granulometrická křivka* množiny  $X$  vzhledem ke granulometrii  $\Psi = \psi_n$ ,  $n \geq 0$ , je zobrazení

$$PS_\Psi(X)(n) = m(\psi_n(X)) - m(\psi_{n-1}(X)), \quad \forall n > 0. \quad (12.65)$$

Posloupnost otevření  $\Psi(X)$ ,  $n \geq 0$ , je klesající, tj.  $(\psi_0(X) \supseteq \psi_1(X) \supseteq \psi_2(X) \supseteq \dots)$ . Lze tedy zavést granulometrii a granulometrickou křivku.

Nyní granulometrickou analýzu použijme se třídou otevření pro binární obrazy. Výchozí binární obraz se nejdříve převede na šedotónový obraz pomocí granulometrické funkce  $G_\Psi(X)$  a granulometrické spektrum  $PS_\Psi$  se vypočte jako histogram granulometrické funkce.

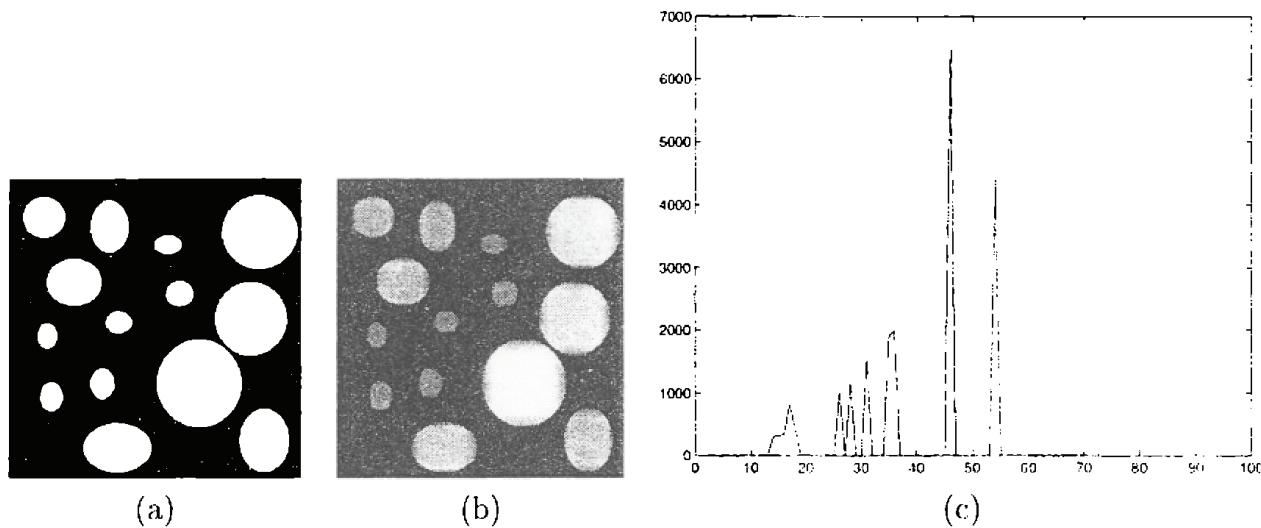
Granulometrická funkce  $G_\Psi(X)$  binárního obrazu  $X$  s využitím granulometrie  $\Psi = (\psi_n)$ ,  $n \geq 0$ , zobrazuje každý pixel  $x \in X$  do velikosti prvního  $n$ , pro které  $x \notin \psi_n(X)$

$$x \in X, G_\Psi(X)(x) = \min\{n > 0, x \notin \psi_n(X)\}. \quad (12.66)$$

Granulometrické spektrum  $PS_\Psi$  binárního obrazu  $X$  pro granulometrii  $\Psi = (\psi_n)$ ,  $n \geq 0$ , lze vypočítat z granulometrické funkce  $G_\Psi(X)$  jako její histogram

$$\forall n > 0, PS_\Psi(X)(n) = \text{card}\{p, G_\Psi(X)(p) = n\}, \quad (12.67)$$

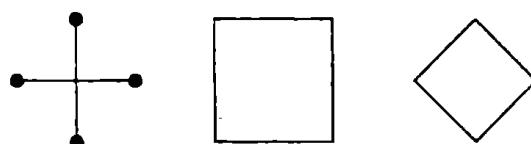
kde  $\text{card}$  označuje kardinalitu množiny. Příklad granulometrické analýzy je na obr. 12.42. Vstupní binární obraz s kruhy různého poloměru je na obr. 12.42a. Obr. 12.42b ukazuje jedno z otevření čtvercovým strukturním elementem. Na obr. 12.42c je granulometrické spektrum. Při záběžném pohledu jsou na granulometrickém spektru patrné tři významnější špičky, které indikují tři převládající velikosti objektů. Méně významné signály v levé části spektra jsou artefakty způsobené diskretizací, protože euklidovské kruhy se nahrazovaly diskrétními objekty čtverci.



Obrázek 12.42 Chování binární granulometrie. (a) Výchozí binární obraz. (b) Maximální vepsané čtverce, výchozí čtverec byl  $2 \times 2$  pixelů. (c) Granulometrické spektrum jako histogram (b), na vodorovné ose je velikost objektů a na svislé ose je počet pixelů objektu dané velikosti. Poděkování P. Kodl, Výzkumné centrum Rockwell Automation, Praha.

Z předchozího příkladu je vidět, že granulometrie umožňuje odvodit informaci o velikosti objektů bez nutnosti předem najít objekty, tj. bez nutnosti obraz segmentovat.

Až do poloviny 90. let byla granulometrie pomocí otevření příliš pomalá na to, aby byla prakticky užitečná. Nyní ovšem nastal obrat k lepšímu. Zájemce o implementaci odkazujeme na [25, 96]. Cestou k urychlení bylo použití podlouhlých strukturálních elementů a složitějších



Obrázek 12.43 Strukturní elementy používané v binární granulometrii vycházejí z tvarů jako kříž, čtverec, kosočtverec na špičce.

dvojrozměrných, které jsou z nich odvozeny, viz 12.43 jako Minkowského součet dvou 1D strukturálních elementů. Například čtvercový strukturní element lze vyjádřit jako Minkowského součet horizontálního a vertikálního čárového elementu.

*Šedotónová granulometrická analýza* patří k nově rozvíjeným nástrojům. Více informací lze najít v [95].

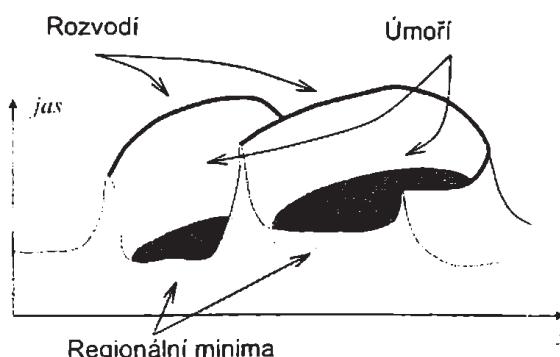
## 12.6 Morfologická segmentace a rozvodí

### 12.6.1 Segmentace častic, jejich značkování a rozvodí

Pojem "segmentace" v analýze obrazu znamená nalezení objektů v obrazu se známou interpretací, které nás zajímají. V tomto skriptu jsme o segmentaci záměrně nemluvili, protože segmentace zasahuje mimo metody zpracování obrazů jako signálů. V případě morfologické segmentace uděláme výjimku proto, aby výklad o matematické morfologii byl ucelený. Metody matematické morfologie především napomáhají segmentovat obrazy častic jak pro binární, tak pro šedotónové vstupní obrazy. V případě binárních obrazů je segmentační úloha formulována jako oddělování překrývajících se častic a v případě šedotónových obrazů je cílem úlohy najít hranici objektů. Náš výklad je inspirován textem [96].

Morfologická segmentace častic se skládá obvykle ze dvou kroků: (1) nalezení značky identifikující částici, (2) transformace rozvodí rekonstruuje částici ze značky.

Značkování připomíná činnost, kterou očekáváme od člověka, znalce, kdy myší označí bod objektu a nestanovuje jeho hranici. *Značka* objektu nebo množiny  $X$  je množina  $M \subseteq X$ . Značky  $M$  zachovávají homotopii množiny  $X$  a jsou obvykle umístěny v prostřední části objektu (částice).



Obrázek 12.44 Znázornění úmoří a rozvodí při topografickém pohledu na obraz jako na krajinu.

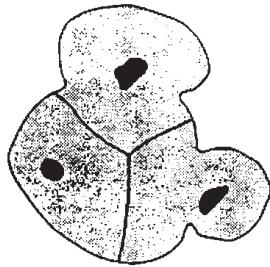
Umístění značek vyžaduje rozumět objektům, jejich sémantice a je proto závislé na konkrétní aplikační oblasti. Většinou se kombinuje nemorfologický a morfologický přístup. Pro komplikované objekty se navíc značkování často přenechá odborníkovi, který značky umístí ručně. Například dobrý software pro analýzu mikroskopických obrazů obsahuje uživatelsky příjemné nástroje pro ruční a semiautomatické značkování.

Když jsou objekty označkovány, mohou být jim odpovídající oblasti nalezeny narůstáním ze značek pomocí transformace rozvodí, což uvidíme v části 12.6.3). Zde budeme vycházet z topografické analogie. Představme si krajinu při dešti. Voda najde klesání a dostane se do jezera nebo moře. Víme již, že jezera a moře odpovídají regionálním minimům. Krajinu lze rozložit na oblasti, do kterých se dešťová voda stahuje. Oblast, odkud voda odtéká do určitého jezera nebo moře, nazývejme pro jednoduchost úmořím. Můžeme také říci, že úmoří jsou oblastní vlivu

regionálních minim obrazové funkce. *Rozvodí* (angl. watershed), nazývané také *linie rozvodí* oddělují jednotlivá úmoří. Úmoří a rozvodí jsou znázorněna na obr. 12.44.

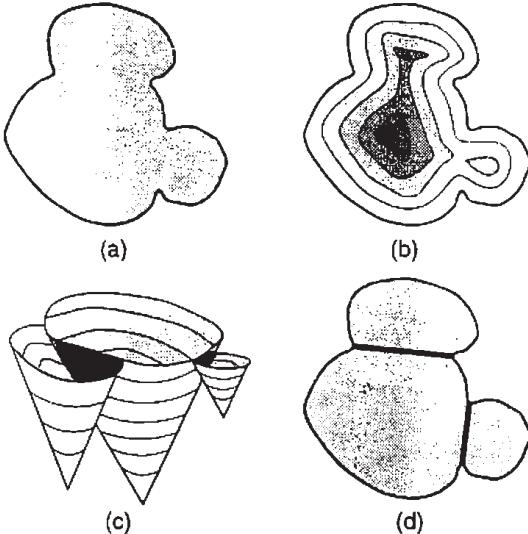
### 12.6.2 Morfologická segmentace binárních obrazů

Když je úkolem najít objekty lišící se jasem od pozadí s proměnlivým jasem, je možné použít transformaci *vrchní část klobouku*, kterou jsme vysvětlili v části 12.4.4. Vrchní část klobouku najde jasová převýšení (špičky), kterými se obraz odlišuje od lokálního pozadí. Při transformaci nemá žádný vliv jasový průběh obrazové funkce, závisí jen na tvaru definičního oboru strukturního elementu. Transformace rozvodí bere v úvahu obojí.



Obrázek 12.45 Segmentace pomocí geodetických zón vlivu nevede ke správným výsledkům.

Morfologická segmentace binárních obrazů nalézá oblasti, které odpovídají jednotlivým překrývajícím se objektům (většinou částicí). Většinu potřebných nástrojů jsme si již vysvětlili. Pro označkování může být použita konečná eroze, jak jsme viděli v části 12.4.4, nebo se značky umístí ručně. Nyní započne narůstání ze značek, narůstaná oblast nemá přesáhnout výchozí množiny a oblasti se nemají spojit.



Obrázek 12.46 Segmentace částic v binárním obrazu. (a) Výchozí obraz. (b) Šedotónový obraz vznikne z (a) pomocí vzdálenostní funkce. (c) Topografický pohled na úmoří. (d) Správně segmentované částice jako rozvodí obrazu (b).

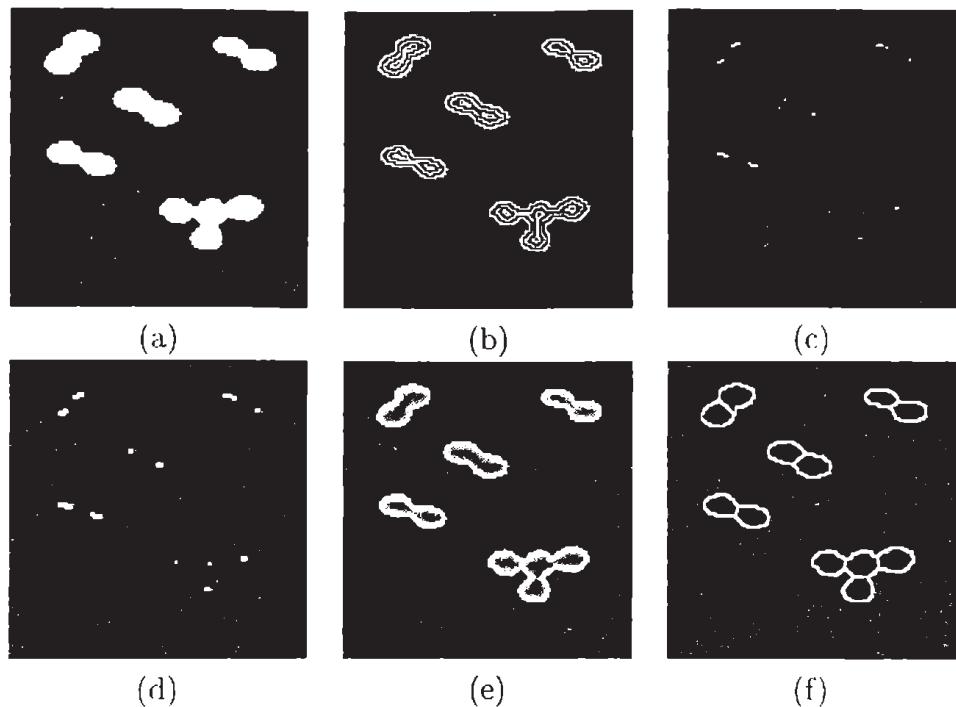
Starší postup pro tento účel se nazývá *podmíněná dilatace*. Pro růst ze značky je použita dilatace a výsledek je dodatečně omezen dvěma podmínkami, tj. zůstat uvnitř původních množin a nespojit částice.

*Geodetická rekonstrukce* probraná v části 12.4.7) je důmyslnější a mnohem rychlejší než podmíněná dilatace. Strukturní element se přizpůsobuje podle hodnot v okolí zpracovávaného

pixelu.

*Geodetické zóny vlivu* z části 12.4.7 se také někdy používají pro segmentaci částic. Obr. 12.45 ukazuje, že výsledek se může lišit od toho, co bychom intuitivně očekávali.

Nejlepším řešením je transformace *rozvodí*. Naznačme princip z hlediska použití pro segmentaci částic. K algoritmické stránce se dostaneme v příští části 12.6.3. Výchozí binární obraz se převede na šedotónový pomocí záporně vzaté vzdálenostní transformace  $-dist$ , viz rovnice (12.47). Pokud kapka deště spadne na topograficky chápáný obraz  $-dist$ , odteče největším spádem do regionálního minima. Myšlenku má ilustrovat obr. 12.46c,d.



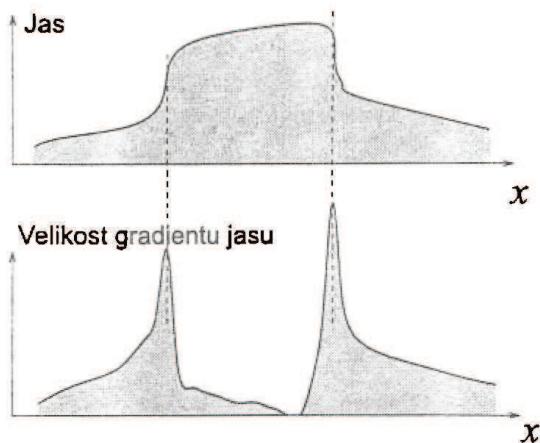
**Obrázek 12.47** Segmentace částic pomocí rozvodí. (a) Výchozí binární obraz. (b) Vzdálenostní funkce zobrazená pomocí vrstevnic. (c) Za značky jednotlivých částic se vezmou regionální maxima vzdálenostní funkce. (d) Dilatované značky. (e) Záporně vzatá vzdálenostní funkce s přidanými značkami. (f) Výsledné hranice částic získané segmentací pomocí rozvodí. Poděkování P. Kodl, Výzkumné centrum Rockwell Automation, Praha.

Použití transformace rozvodí na segmentaci částic si ukažme na příkladě na obr. 12.47. Vstupem je binární obraz několika překrývajících se částic, obr. 12.47(a). Značkami označujícími jednotlivé částice jsou regionální maxima vzdálenostní funkce 12.47(c). Značky se dilatují, jak je vidět z obr. 12.47(d). Přípravou pro segmentaci pomocí rozvodí je obrácení znaménka u vzdálenostní transformace. Výsledek spolu s dilatovanými značkami je na obr. 12.47(e). Výsledek, tj. jednotlivé částice a jejich hranice, je na obr. 12.47(f).

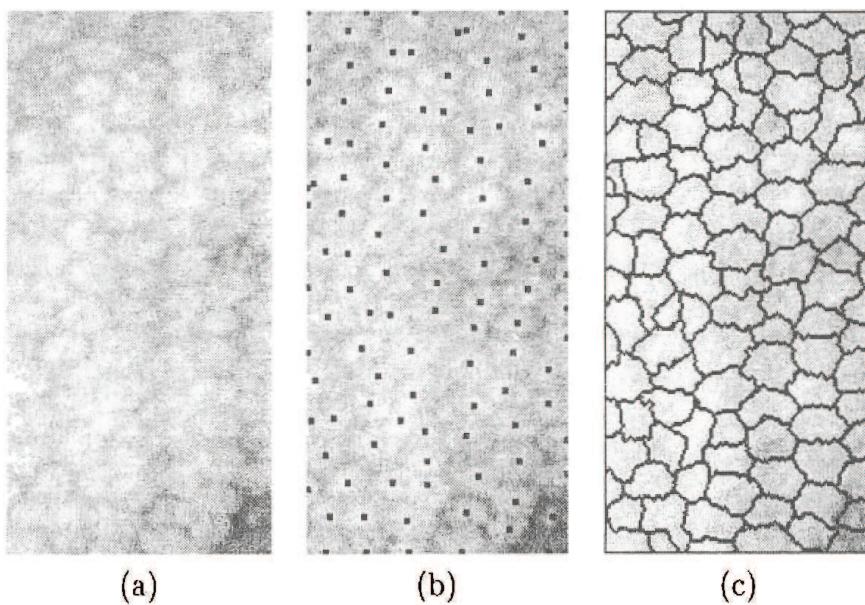
### 12.6.3 Šedotónová segmentace, rozvodí

Rozvodí [94, 96] lze také využít pro hledání hranic objektů v šedotónových obrazech, odpovídajících nejrychlejším změnám velikosti obrazové funkce. Transformace rozvodí se použije na gradientní obraz vstupního obrazu, což je ilustrováno na obr. 12.48. Existuje výpočetně jednoduchá approximace gradientu obrazu zvaná *Beucherův gradient* [87], počítaný jako množinový rozdíl mezi dilatací vstupního obrazu  $X$  jednotkovým kruhem a erozí  $X$  jednotkovým kruhem

$$grad(X) = (X \oplus B) - (X \ominus B). \quad (12.68)$$



**Obrázek 12.48** Segmentace v šedotónových obrazech pomocí gradientu.



**Obrázek 12.49** Segmentace šedotónového obrazu sítnice lidského oka. (a) Výchozí šedotónový obraz. (b) Body označují značky, které byly nalezeny nemorfologickou metodou. (c) Hranice buněk na sítnici nalezené jako rozvodí ze značek (b). Poděkování za data a značky R. Šára, Elektrotechnická fakulta ČVUT Praha, segmentace P. Kodl, Výzkumné centrum Rockwell Automation Praha.

Klasické metody segmentace počítané přes gradient obrazové funkce trpí *přesegmentováním*, což znamená, že obraz je rozdělen do příliš mnoha oblastí (12.47c). Některé morfologické postupy tomuto jevu dokážou zabránit [94].

Ukažme si na příkladě z oftalmologie použití segmentace přes rozvodí. Vstupním obrazem je pohled na část sítnice lidského oka, obr. 12.49.

# Literatura

- [1] R. H. T. Bates a M. J. McJannett. *Image Restoration and Reconstruction*. Clarendon Press, Oxford, England, 1980.
- [2] J.S. Bendat a A.G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. J.Wiley and Sons, 1970.
- [3] H. Blum. A transformation for extracting new descriptors of shape. *Proceedings of the Symposium on models for the perception of Speech and Visual Form*, red. W. Wathen-Dunn, Cambridge, MA., USA, November 1964. s. 362–380.
- [4] J. Canny. A computational approach to edge detection. *IEEE on PAMI*, 8(6):679–698, 1986.
- [5] V. Čížek. *Diskrétní Fourierova transformace a její použití*. SNTL, Praha, 1981.
- [6] J.W. Cooley, Lewis. P.A a P.D. Welch. *Historical Notes on the Fast Fourier transforms*. IEEE Press, 1970. in Rabiner, L. and Rader, Ch.M.(ed.): *Digital Signal Processing*.
- [7] J.W. Cooley, Lewis. P.A a P.D. Welch. *Fast Fourier transform and Its Use for Time-Series Analysis*. J.Wiley and Sons, 1977. in *Statistical Methods for Digital Computers*.
- [8] J.W. Cooley a J.W. Tukey. An algorithm for machine calculation of complex fourier series. *Math. Computation*, 19:297–301, 1965.
- [9] S. Ďádó a M. Kreidl. *Senzory a měřicí obvody*. Vydavatelství ČVUT, Praha, 1996.
- [10] I. Daubechies. Ten lectures on wavelets. In: *SIAM*, 1992.
- [11] V. Davídek a P. Sovka. *Digitové zpracování signálů a implementace*. Vydavatelství ČVUT, Praha, 1996.
- [12] Analog Devices. *Digital signal processing Applications Using the ADSP-2100 Family*. Prentice Hall, N.J., 1992.
- [13] E. R. Dougherty. *An introduction to Morphological Image Processing*. SPIE Press, Bellingham, WA, 1992.
- [14] D. E. Dudgeon a R. M. Mersereau. *Multidimensional signal processing*. Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1984. Russian translation Mir, Moskva 1988, 488 p.
- [15] A.W.M. Enden a N.A.M. Verhoeck. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [16] J. D. Foley, A. van Dam, S. K. Feiner a J. F. Hughes. *Computer Graphics - Principles and Practice*. Addison-Wesley Publishing Company, second vydání, 1993.
- [17] C. R. Giardina a E. R. Dougherty. *Morphological Methods in Image and Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [18] M. J. E. Golay. Hexagonal parallel pattern transformation. *IEEE Transactions on Computers*, C-18:733–740, 1969.
- [19] R. C. Gonzalez a R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, Ma, 1992.
- [20] J. Goswami a A. Chan. *Fundamentals of Wavelets: Theory, Algorithms and Applications*. John Wiley Interscience, New York, N.Y., 1999.
- [21] Gösta H. Granlund a Hans Knutss. *Signal processing for computer vision*. Kluwer Academic Publisher, Dordrecht, The Netherlands, 1995.

- [22] A. Grays. An introduction to wavelets. *IEEE Comp. Science and Engineering*, 2(2), 1995.
- [23] V. Haasz a M. Sedláček. *Elektrická měření. Přístroje a metody*. Vydavatelství ČVUT, Praha, 1998.
- [24] R. M. Haralick. Performance characterization in computer vision. *Proceedings of the 5th International Conference on Computer Analysis of Images and Patterns*, red. D. Chetverikov a W. G. Kropatsch, Heidelberg, September 1993. Springer-Verlag Heidelberg, Lecture Notes in Computer Science No. 719, s. 1–9.
- [25] R. M. Haralick, P. L. Katz a E. R. Dougherty. Model-based morphology: the opening spectrum. *Graphical Models and Image Processing*, 57(1):1–12, 1995.
- [26] R. M. Haralick a L. G. Shapiro. *Computer and Robot Vision, Volume I*. Addison Wesley, Reading, Ma., 1992.
- [27] R. M. Haralick a L. G. Shapiro. *Computer and Robot Vision, Volume II*. Addison Wesley, Reading, Ma., 1993.
- [28] R. M. Haralick, S. R. Stenberg a X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532–550, 1987.
- [29] H. J. A. M. Heijmans. *Morphological Image Operators*. Academic Press, Boston, MA., 1994.
- [30] R.J. Higgins. *Digital Signal Processing in VLSI*. Prentice Hall, Englewood Cliffs, 1990. with Analog Devices.
- [31] V. Hlaváč a M. Šonka. Zpracování obrazu počítačem. skriptum FEL ČVUT, Ediční středisko ČVUT, 1991.
- [32] V. Hlaváč a M. Šonka. *Počítačové vidění*. Grada Praha, 1992.
- [33] Z. Horák a F. Krupka. *Fyzika, příručka pro fakulty strojního inženýrství*. SNTL, Praha, 1966.
- [34] B. K. P. Horn. *Robot Vision*. The MIT Press, Cambridge, Massachusetts, USA, 1986.
- [35] B. K. P. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75, 1990.
- [36] R. E. Hufnagel a N. R. Stanley. Modulation transfer function associated with image transmission through turbulent media. *Journal of the Optical Society of America*, 54:52–61, 1964.
- [37] K. Ikeuchi. Surface reflection mechanism. svazek 2, San Diego, USA, 1994. Academic Press, s. 131–160.
- [38] K. Ikeuchi a B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.
- [39] V.K. Ingle a J.G. Proakis. *Digital Signal Processing Laboratory Using the ADSP-2101 Microcomputer*. Prentice Hall, N.J., 1991.
- [40] Texas Instruments. *Digital Signal Processing Applications with TMS 320 Family. Theory, Algorithms and Implementations*. Texas Instruments, 1990.
- [41] Bernd Jähne. *Practical Handbook on Image Processing for Scientific Applications*. CRC Press, Boca Raton, Florida, 1997.
- [42] Anil K. Jain. *Fundamental of Digital Image Processing*. Prentice Hall Information and System Science Series. Prentice Hall, 1989.
- [43] J. Jan. *Číslicová filtrace, analýza a restaurace signálů*. Vysoké učení technické, Brno, 1997.
- [44] L. P. Jaroslavskij a I. Bajla. *Metódy a systémy číslicového spracovania obrazov*. Alfa, Bratislava, 1989.

- [45] K.D. Kammeyer a K. Kroschel. *Digitale Signalverarbeitung. Filterung und Spektralanalyse*. Tuebner Studienbücher, B.G.Teubner, 1989.
- [46] S. M. Kay a S. L. jr. Marple. Spectrum analysis - a modern perspective. *Proc. IEEE*, 69(11):pp. 1380–1419, Nov. 1981.
- [47] R. Klette, A. Koschan a K. Schlüns. *Computer Vision - Räumliche Information aus digitalen Bildern*. Friedr. Vieweg & Sohn, Braunschweig, 1996.
- [48] Miloš Klíma, Martin Bernas, Jiří Hozman a Pavel Dvořák. *Zpracování obrazové informace*. Vydavatelství ČVUT, 1996.
- [49] V. Kovalevski. Digital geometry based on the topology of abstract cell complexes. In: *Proceedings of the Colloque Géometrie Discrete En Imagerie*, Strasbourg, September 1993. s. 259–284.
- [50] V. A. Kovalevsky. Digital geometry based on the topology of abstract cell complexes. *Géometrie Discrète en Imagery, Fondements et Applications*, red. Jean-Marc Chassery, Jean Françon, Annick Montanvert a Jean-Pierre Réveillès, Strasbourg, září 1993. s. 259–283.
- [51] W. Kropatsch a P. Pták. The path-connectedness in  $Z^2$  and  $Z^3$  and classical topologies (the point-neighbourhood formalism). *Advances in Pattern Recognition, Proceedings of Joint IAPR International Workshop SSPR'98 and SPR'98*, red. A. Amin, D. Dori, P. Pudil a H. Freeman, číslo 1451 v Lecture Notes in Computer Science, Berlin, Germany, August 1998. Springer-Verlag, s. 181–189.
- [52] H. Y. F. Lam. *Analog and Digital Filters: Design nad Realisation*. Prentice Hall, Inc., Englewood Cliffs, N.Y., 1979.
- [53] B. R. Levin. *Teorie náhodných procesů a její aplikace v radiotechnice*. SNTL, Praha, 1965. (Překlad z ruštiny).
- [54] J. Likeš a J. Machek. *Počet pravděpodobnosti*. SNTL, Praha, 1982.
- [55] J. Likeš a J. Machek. *Matematická statistika*. SNTL, Praha, 1983.
- [56] P. Maragos a R. W. Schafer. Morphological filters – part I: Their set-theoretic analysis and relations to linear shift-invariant filters. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(8):1153–1169, August 1987.
- [57] P. Maragos a R. W. Schafer. Morphological filters – part II: Their relation to median, order-statistic, and stack filters. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(8):1170–1184, August 1987.
- [58] P. Maragos a R. Ziff. Threshold superposition in morphological image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), May 1990.
- [59] D. Marr. *Vision – A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Co., San Francisco, 1982.
- [60] G. Matheron. *Eléments pour une Theorie des Milieux Poreux (in French)*. Masson, Paris, 1967.
- [61] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, New York, 1975.
- [62] V. Matyáš. *Měření, analýza a vytváření náhodných procesů*. SNTL, Praha, 1976.
- [63] V. Mařík a al. *Teorie Automatického řízení II, Příklady*. ES ČVUT, 1989. Skripta ČVUT.
- [64] F. Meyer. Contrast feature extraction. *Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, red. J.-L. Chermant, Stuttgart, Germany, 1978. Riederer Verlag. Special issue of Practical Metallography.
- [65] Yves Meyer. *Wavelets: Algorithms and Applications*. Society for Industrial and Applied Mathematics, 1993.

- [66] F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg a T. Limperis. Geometrical considerations and nomenclature for reflectance. U.S. Department of Commerce, National Bureau of Standards, Washington D.C., USA, 1977.
- [67] J. Oliensis. Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 54(2):163–183, September 1991.
- [68] A. V. Oppenheim a R.W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [69] D. Orofino. Wavelets in real-time applications. *MATLAB News and Notes*, Summer 1999:pp. 12–13, 1999.
- [70] A. Papoulis. *The Fourier Integral and Its Application*. McGraw Hill, New York, 1962.
- [71] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, New York, 1965.
- [72] A. Papoulis. *Signal analysis*. McGraw-Hill, 1977.
- [73] T. W. Parks a J. M. McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circuit Theory*, CT-19, March 1972.
- [74] T. W. Parks a J. M. McClellan. A program for the design of linear phase finite impulse response filters. *IEEE Trans. Audio Electroacoust.*, AU-20(3), August 1972.
- [75] T. Pavlidis. *Structural Pattern Recognition*. Springer-Verlag, Berlin, 1977.
- [76] Z. Pírko a J. Veit. *Laplaceova transformace*. SNTL, Praha, 1970.
- [77] W. K. Pratt. *Digital Image Processing*. John Wiley and Sons, New York, 2nd vydání, 1991.
- [78] R. B. Randall. *Frequency Analysis*. Brüel and Kjaer, 3rd vydání, 1987.
- [79] K. R. Rao a P. Yip. *Discrete Cosine Transform, Algorithms, Advantages, Applications*. Academic Press, Boston, Ma, 1990.
- [80] L. G. Roberts. Machine perception of three-dimensional solids. *Optical and Electro-Optical Information Processing*, red. J. T. Tippett. MIT Press, Cambridge, Ma, 1965, s. 159–197.
- [81] A. Rosenfeld a A. C. Kak. *Digital Picture Processing*. Academic Press, New York, 2nd vydání, 1982.
- [82] E. Ružický a A. Ferkov. *Počítačová grafika a spracovanie obrazu*. Sapientia Bratislava, 1995.
- [83] Radim Šára. *Local Shading Analysis via Isophotes Properties*. Dr, Johannes Kepler University Linz, Dept. of System Sciences, Linz, Austria, March 1994.
- [84] Ed. Schnell, L. *Technology of Electrical Measurements*. John Wiley and Sons, Chichester, 1993.
- [85] E. Schröfer. *Signalverarbeitung. Numerische Verarbeitung digitaler Signale*. Carl Hauser Verlag, München - Wien, 1990.
- [86] M. Sedláček. *Zpracování signálů v měřicí technice*. Vydavatelství ČVUT, Praha, 1998.
- [87] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.
- [88] J. Serra. Morphological optics. *Journal of Microscopy*, 145(1):1–22, 1987.
- [89] M. Šonka, V. Hlaváč a R.D. Boyle. *Image Processing, Analysis and Machine Vision*. PWS, Boston, USA, second vydání, 1998.
- [90] L. Špaček. Edge detection and motion detection. *Image and Vision Computing*, 4:43–52, 1986.
- [91] Jan Uhlíř a Pavel Sovka. *Číslicové zpracování signálů*. Vydavatelství ČVUT, 1995.
- [92] R. Vích. *Transformace Z a některá její použití*. SNTL, 1979.
- [93] L. Vincent. Efficient computation of various types of skeletons. In: *Proceedings of the SPIE Symposium Medical Imaging V*, Vol. 1445, San Jose, CA. SPIE, February 1991.

- [94] L. Vincent. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2(2):176–201, April 1993.
- [95] L. Vincent. Fast opening functions and morphological granulometries. In: *Proceedings Image Algebra and Morphological Image Processing V, San Diego, CA*. SPIE, July 1994, s. 253–267.
- [96] L. Vincent. Lecture notes on granulometries, segmentation and morphological algorithms. *Proceedings of the Summer School on Morphological Image and Signal Processing, Zakopane, Poland*, red. K. Wojciechowski. Silesian Technical University, Gliwice, Poland, September 1995, s. 119–216.
- [97] R. Weiss. 32-bit floating-point dsp processors. *EDN Special Report*, November 1991.
- [98] P.D. Welch. The use of fast fourier transform for the estimation of power spectra. *IEEE Trans. Audio Electroacoustics*, AU-15:70–73, June 1970.
- [99] A. P. Witkin. Scale space filtering. *From Pixels to Predicates*, red. A. P. Pentland. Ablex Publishing Corporation, Norwood, NJ, 1986, s. 5–19.
- [100] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.
- [101] Jiří Žára, Bedřich Beneš a Petr Felkel. *Moderní počítačová grafika*. Computer Press, Praha, 1998.

# Rejstřík

**A**daptivní filtr, 56

AGC, 134

albedo, 125

aliasing, 8, 35

analýzátor

— signálu, 36

— spektrální, 36

apertura, 131

aproximace jasové funkce, 138

azimut, 124

**B**DRF, 125

bias, 80

bod

— reprezentativní, 178

— vlícovací, 137

bodová množina, 177

**C**esta, 116

chyba

— rekonstrukce signálu, 12

— středně kvadratická, 80

— useknutím, 35

clona, 131

cyklus mezní, 45

**D**ata obrazová, 108

derivátor, 44, 51

DFT, 21

— motýlek, 34

— mřížka, 22

— přímá, 22

— zpětná, 22

digitalizace, 114

dilatace

— binární, 179

— geodetická, 200

— podmíněná, 206

— šedotónová, 187

Diracův impuls, 5

diskrétní topologie, 118

disperse, 82

distribuční funkce, 76

dolní propust, 50

— Besselova, 68

— Butterworthova, 67

— Cauerova, 68

— eliptická, 68

— Čebyševova, 67

dualita, 178

díra, 117

dírková komora, 112, 130

**E**lement strukturní, 178

erosion, 183

eroze

— binární, 181

— geodetická, 200

— konečná, 198

— šedotónová, 187

**F**FT, 27, 31

filtr

— adaptivní, 56

— antialiasingový, 13

— autokorelační, 90

— Besselův, 68

— Butterworthův, 67

— Cauerův, 68

— derivátor, 51

— dolní propust, 153

— eliptický, 68

— FIR, 41

— frekvenčně selektivní, 50

— frekvenční, 105

— fázový posouvač, 53

— gaussovský, 169

— Hilbertův transformátor, 52

- homogenní, 159
  - horní propust, 153
  - IIR, 42
    - IIR kaskádní forma, 48
    - IIR paralelní forma, 48
    - IIR přímá forma, 46
  - integrátor, 50
  - kanonická forma, 42
  - klouzavý průměr, 53, 120
  - korelační, 90, 103
  - lineární, 153
  - nelineární, 161
  - nerekurzivní, 42
  - návrh, 57
  - návrh pomocí prototypu, 66
  - pomocí vzájemné korelační funkce, 92
  - prostorově invariantní, 159
  - pásmová propust, 153
  - přenos, 46
  - RDF, 44
  - rekonstrukční, 12
  - rekurzivní, 42, 44
    - s konečnou impulsní odezvou, 41
    - s lineární fází, 44
    - s nekonečnou impulsní odezvou, 42
  - separabilní, 160
  - transpozice, 72
  - v prostorové oblasti, 145
    - ve frekvenční oblasti, 145
  - Čebyševův, 67
  - číslicový, 41
  - řád, 45
  - filtrace
    - inverzní, 175
    - lokální, 157
    - Wienerova, 175
  - fotometrické stereo, 128
  - fotometrie, 122
  - Fourierova transformace
    - diskrétní, 16, 21
    - rychlá, 27, 31
    - vlastnosti, 18
  - frekvence
    - normovaná kruhová, 17
    - Nyquistova, 5
  - frekvenční bin, 22
  - frekvenční spektrum, 16
  - FTD, 16
- funkce
    - autokorelační, 85
    - autokovarianční, 85
    - delta, 5
    - distribuční, 76
    - granulometrická, 203
    - Haarova, 151
    - kauzální mocninná, 21
    - konjugovaně sudá, 18
    - korelační cyklická, 94
    - matečná, 151
    - obrazová, 108, 112
    - odrazivosti, 125
    - periodické prodloužení, 23
    - vzdálenostní, 198
    - vzájemná korelační, 85
    - vzájemná kovarianční, 85
  - fázový posouvač, 53
- G**ama korekce, 134  
geodetická transformace, 199  
globální extrém, 196  
Golayova abeceda, 193  
gradient, 119, 164
  - Beucherův, 207

**H**istogram, 79, 119
  - ekvalizace, 143
  - kumulativní, 144
  - vyrovnání, 143

**hlavní bod**, 130  
**hlavní lalok**, 37  
**hloubka**
  - ostrosti, 131
  - pole, 132

**hodnota**
  - očekávaná, 81
  - střední, 81

**homogenní souřadnice**, 137  
**homotopická transformace**, 190  
**horní propust**, 50  
**hrana**, 119  
**hranice oblasti**, 118
  - vnitřní, 118
  - vnější, 118

**hustota**

— pravděpodobnosti, 76  
 — výkonová spektrální, 148

## I

### Identifikace, 92

impuls  
 — delta, 5  
 — Diracův, 5  
 — jednotkový, 5, 14  
 — obdélníkový, 19

integrátor, 50  
 intenzitní obraz, 108  
 interpolace, 138

— Shannonova, 9

interpolace jasu  
 — bikubická, 140  
 — lineární, 139  
 — nejbližší soused, 139

interpolátor  
 — boxcar, 10  
 — Lagrangeův, 10  
 — lineární, 11  
 — sample-and-hold, 10  
 — zero-order hold, 10

interpretace, 109

interval vzorkovací, 4

## J

### Jas, 108

— definice, 122

jasové korekce, 141

jezero, 118

jitter, 36, 134

## K

### Kamera, 130, 134, 174

— analogová, 134  
 — pixlově synchronizovaná, 134

klasifikace, 109

klouzavý průměr, 53

koefficient odrazivosti, 125

komplexní exponenciála, 15

konečná eroze, 198

konformní zobrazení, 69

konvexní obal, 118

konvoluce, 18, 21, 148, 158

— cyklická, 29

— diskrétní aperiodická, 29

— kruhová, 29

— lineární, 29

— rychlá, 31

— segmentovaná, 31

konvoluční maska, 158

korekce

— gama, 134

— gamma, 142

— jasové, 141

kostra množiny, 191

kruh

— geodetický, 199

— maximální, 191

kvantování, 114

křivka granulometrická, 203

## L

### Lalok, 37

lambertovský povrch, 126

Laplacián, 164

leakage, 26, 37, 105

lokální extrém, 196

## M

### Mapa odrazivosti, 126

matice

— hermitovská, 146

— ortogonální, 146

— symetrická, 146

— unitární, 146

maximální kruh, 191

medián, 83

metoda

— Blackmanova-Tuckeyova, 101

— Welchova, 100

mezní cyklus, 45

množina

— bodová, 177

— symetrická, 182

— transponovaná, 182

množinový rozdíl, 177

model, 108

modus, 83

moment, 80

— centrovány, 81

— centrální, 81

— obecný, 81

měrná svítivost, 122

měřítko, 170

mřížka

— DFT, 22

mřížka, 114

## O

### Objekt, 110

- objektiv
    - *f*-číslo, 124
    - rozostřený, 175
    - telecentrický, 133
  - oblast, 116
  - obraz, 108
    - barevný, 120
    - digitalizace, 114
    - gradient, 119
    - intenzitní, 108
    - kvantování, 114
    - monochromatický, 120
    - multispektrální, 120
    - obnovení, 173
    - ostření, 164
    - paletový, 142
    - pseudobarevný, 142
    - reprezentace, 110
    - restaurace, 173
  - obrazová funkce, 108
  - odezva impulsní, 6
  - odhad
    - hustoty pravděpodobnosti, 78
    - konzistentní, 80
    - vychýlení, 80
  - ohnisko, 131
  - okno, 27
    - Barlett, 61
    - Blackman, 61
    - Dolph-Čebyšev, 64
    - Hamming, 61
    - Hann, 27, 38, 61
    - Kaiser, 64
    - obdélníkové, 28, 61
    - trojúhelníkové, 61
    - časové, 27
  - operace lokální, 119
  - operátor
    - Cannyho, 171
    - gradientní, 164
    - Kirschův, 167
    - Laplaceův, 164, 166
    - LoG, 169
    - Marrův-Hildrethové, 168
    - Prewittové, 167
    - Robertsův, 166
    - Robinsonův, 167
    - Sobelův, 167
  - ostření obrazu, 164
  - osvětlení, 122
  - otevření
    - binární, 183
    - šedotónové, 189
  - otáčecí činitel, 32
  - ozáření, 122
- P**erioda vzorkovací, 4
- periodogram, 96
  - pixel, 114
    - sousednost, 116
    - souvislý, 116
  - polární úhel, 124
  - populace, 157
  - porucha
    - rozmazání polybem, 174
    - rozostřený objektiv, 175
    - turbulence atmosféry, 175
  - posloupnost, 3, 15
  - segmentace, 31
  - postranní lalok, 37
  - povrch
    - ideálně difuzní, 126
    - ideálně matný, 126
    - lambertovský, 126
  - pozadí, 117
  - počítačová grafika, 109
  - počítačové vidění, 108
  - proces
    - náhodný, 74
    - náhodný ergodický, 88
    - stochastický, 74
  - procesor signálový, 41
  - propust
    - dolní, 44, 50
    - horní, 44, 50
    - pásmová, 44, 50
  - prostor
    - gradientní, 126
  - prostor měřítek, 171
  - prostorový úhel, 122
  - prosívání, 203
  - průměr
    - výběrový, 157
  - průměrování
    - klouzavým průměrem, 105
    - rekurzivní, 104

- signálu, 104
- pásmová propust, 50
- pásmová zádrž, 50
- přenos, 46
- příznak, 110

### **Radometrie, 121**

- realizace náhodného procesu, 74

- regionální extrém, 196

- rekonstrukce

- geodetická, 206

- signálu, 5

- signálu ze vzorků, 8

- Whittakerova, 9

- rekonstrukce (morfologická), 198, 200

- repräsentace

- ikonická, 110

- relační model, 110

- robustní statistika, 162

- rovina

- obrazová, 131

- předmětová, 131

- rovnice

- ozáření, 124

- čočky, 130

- rovnice ozáření, 127

- rozdělení

- Gaussovo, 84

- normální, 84

- rovnoměrné, 85

- tvar, 83

- rozlišení

- plošné, 115

- radiometrické, 115

- spektrální, 115

- časové, 115

- rozptyl, 82

- rozvodí, 206

### **Segmentace, 110**

- morfologická, 205

- signál

- analogový, 2

- antiperiodický, 2

- deterministický, 2, 74

- diskrétní v čase, 3

- doplnění nulami, 28

- energetický, 88

- harmonický, 14
- kvaziperiodický, 2
- náhodný, 2, 74
- náhodný ergodický, 75
- náhodný stacionární, 74
- periodicitá, 90

- periodický, 2

- průměrování, 104

- přechodný, 2

- stochastický, 2, 74

- signály rozdělení, 2

- skelet, 191

- homotopický, 192

- maximální kruhy, 192

- pomocí zón vlivu, 199

- SKIZ, 199

- skok jednotkový, 14

- snímač

- CCD, 134

- CMOS, 134

- fotoemisní, 133

- fotovodivostní, 134

- fotovoltaický, 134

- soused, 116

- souřadnice

- homogenní, 137

- projektivní, 137

- spektrum

- frekvenční, 16, 147, 203

- fázové, 147

- granulometrické, 203

- rozmařání, 26, 37

- vlastnosti, 24

- vzorkovaného signálu, 6

- výkonové, 147

- strukturní element, 178

- stupeň šedi, 115, 122

- stín, 186

- střední hodnota, 81

- střední osa, 191

- světelní účinnost, 122

- synchرونní detekce, 106

- Topografický pohled, 185**

- topologie, 177

- transformace

- affinní, 137

- antiextenzivní, 184

- bilineární, 68, 137
  - extenzivní, 184
  - Fourierova, 146
  - Fourierova diskrétní, 21
  - frekvenční, 71
  - geodetická, 199
  - geometrické, 136
  - Haarova, 152
  - Hadamard, 152
  - Hadamard-Haar, 152
  - Hadamardova, 149
  - hodnot jasu, 141
  - homeomorfní, 119
  - homotopická, 190
  - Hough, 152
  - idempotentní, 193
  - jasové stupnice, 141
  - Karhunen-Loeve, 152
  - kosinová, 150
  - lineární integrální, 145
  - ortogonální, 146
  - Paley, 152
  - Radon, 152
  - Slant-Haar, 152
  - souřadnic bodů, 137
  - tref či miň, 183
  - vlnková, 107, 150
  - vrchní část klobouku, 189, 206
  - Walsh, 152
  - Walsh-Hadamard, 149
  - wavelets, 150
  - transformátor Hilbertův, 44, 52
  - translace, 178
  - transpozice filtru, 72
  - turbulence atmosféry, 175
  - tvar rozdělení, 83
  - tvar ze stínování, 127
  - twiddle factor, 32
- Uzavření**
- binární, 184
  - šedotónové, 189
- Variance**, 82
- vinětace, 124
- vlnová délka, 120
- vršek, 186
- vychýlení odhadu, 80
- vyhlazování šumu, 158
  - vzdálenost
    - definice, 115
    - euklidovská, 116
    - geodetická, 199
    - ohnisková, 131
    - v městských blocích, 116
    - šachovnicová, 116
  - vzdálenostní funkce, 198
  - vzorkování, 4
    - adaptivní, 4
    - interval, 114
    - mřížka, 114
    - náhodné, 4
    - náhodných signálů, 80
    - obrazu, 114
    - s pevnou vzorkovací frekvencí, 4
    - sekvenční, 4
  - vzorkování periodické, 4
  - věta
    - centrální limitní, 157
    - o homeomorfismu stínů, 188
    - o transpozici, 72
    - vzorkovací, 4
  - výběr náhodný, 157
  - výběrový soubor, 80
  - výkonová spektrální hustota, 95
- W**atershed, 206
- wavelets, 107, 150
- Welchova metoda, 100
- Wienerův-Chinčinův vztah, 88
- Z**ero padding, 28
- značka, 205
- značka (v morfologii), 198
- zobrazení
  - konformní, 69
  - perspektivní, 109, 112
- zpožděná posloupnost, 15
- zpožďovací člen, 42
- zpracování signálů
  - v reálném čase, 1
- ztenčování, 193
  - sekvenční, 193
- ztlušťování, 193
- ztlušťování sekvenční, 193
- zvětšení optické soustavy, 131

- zádrž — vyhlazování, 158  
— pásmová, 44, 50  
záliv, 118  
zář, 122  
zářivý tok, 122  
**Šum**, 153  
— aditivní, 103, 157  
— bílý, 88  
— bílý frekvenčně omezený, 99  
— bílý vzorkovaný, 99  
— kvantovací, 13  
**Činitel otáčecí**, 32  
člen zpožďovací, 42  
čočka, 123  
— rovnice, 130  
— tenká, 130  
— tlustá, 132  
**Řád filtru**, 45  
**Úmoří**, 205