approach counted test — both failed — only metamorphic — only ground truth

percentage of failing tests vs normalised error threshold