Combinaison des observables en physique des saveurs

Ecole SOS, IPHC Strasbourg, 3 juillet 2008

Jérôme Charles - CPT Marseille

Cours I

- Introduction à la physique des saveurs
- Le projet CKMfitter, historique, objectifs
- Ajustement de données sur un modèle non linéaire
- Le code CKMfitter, présentation et fonctionnalités

Cours II

- Paramètres de nuisance en statistique fréquentiste
- Un exemple partiellement soluble
- Démonstration du code
- Conclusion et perspectives

Introduction à la physique des saveurs

Le Modèle Standard

Le Modèle Standard de la physique des particules décrit les interactions électromagnétiques, fortes et faibles des leptons, quarks, bosons de jauge et boson de Higgs

Le Modèle Standard est construit sur la base du groupe de jauge $SU(3) \times SU(2) \times U(1)$. Le secteur électrofaible présente une phénoménologie très riche et est relié à plusieurs questions fondamentales

- brisure de la symétrie électrofaible
- nature du Higgs
- hiérarchie des masses
- mélange des saveurs et violation de CP

CP : une symétrie fondamentale

C conjugaison de charge, P parité spatiale,
T renversement du temps
C, P, T sont des symétries fondamentales
C et P sont violées maximalement (fermions chiraux)
CP et T sont faiblement violées (10^{-5})
CPT est une symétrie exacte en théorie quantique des champs
C, P et CP sont des ingrédients cruciaux de la théorie du Big-Bang (baryo- et/ou leptogenèse)

CP : une symétrie fondamentale

C conjugaison de charge, P parité spatiale,
T renversement du temps
C, P, T sont des symétries fondamentales
C et P sont violées maximalement (fermions chiraux)
CP et T sont faiblement violées (10^{-5})
CPT est une symétrie exacte en théorie quantique des champs
C, P et CP sont des ingrédients cruciaux de la théorie du Big-Bang (baryo- et/ou leptogenèse)

Découvertes

- 1964 violation de CP indirecte dans le système des kaons neutres
- 1998 violation de T dans le système des kaons neutres
- 1999 violation de CP directe dans le système des kaons neutres
- 2001 violation de CP induite par le mélange dans le système des B
- 2004 violation de CP directe dans le système des B

CP : une symétrie fondamentale

C conjugaison de charge, P parité spatiale,
T renversement du temps
C, P, T sont des symétries fondamentales
C et P sont violées maximalement (fermions chiraux)
CP et T sont faiblement violées (10^{-5})
CPT est une symétrie exacte en théorie quantique des champs
C, P et CP sont des ingrédients cruciaux de la théorie du Big-Bang (baryo- et/ou leptogenèse)

la violation de CP est très bien décrite par le Modèle Standard, mais nous n'en n'avons toujours pas d'explication dynamique

Découvertes

- 1964 violation de CP indirecte dans le système des kaons neutres
- 1998 violation de T dans le système des kaons neutres
- 1999 violation de CP directe dans le système des kaons neutres
- 2001 violation de CP induite par le mélange dans le système des B
- 2004 violation de CP directe dans le système des B

Mélange des quarks

dans le Modèle Standard les saveurs de quarks sont mélangées par l'interaction faible

$\longrightarrow$ bi-diagonalisation sur la base des états propres *via* la matrice de Cabibbo-Kobayashi-Maskawa (CKM)

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

Mélange des quarks

dans le Modèle Standard les saveurs de quarks sont mélangées par l'interaction faible

$\longrightarrow$ bi-diagonalisation sur la base des états propres *via* la matrice de Cabibbo-Kobayashi-Maskawa (CKM)

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

cette matrice unitaire est complexe ($V_{ub} \propto |V_{ub}|\, e^{-i\gamma}$) dès qu'il y a au moins trois générations de fermions non dégénérés

CKM génère (de) la violation de CP

Hiérarchie et triangles d'unitarité

la matrice CKM présente une forte hiérarchie, d'origine inconnue:

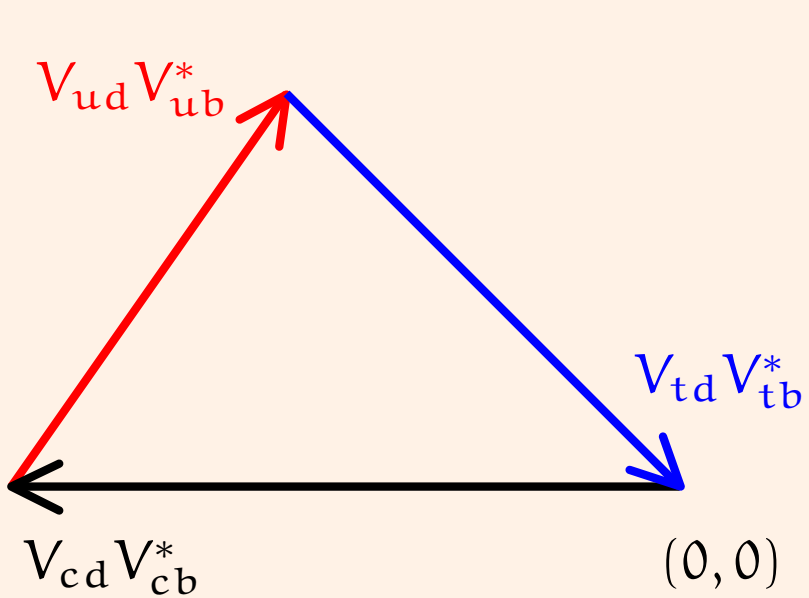
couplages diagonaux $\propto 1$

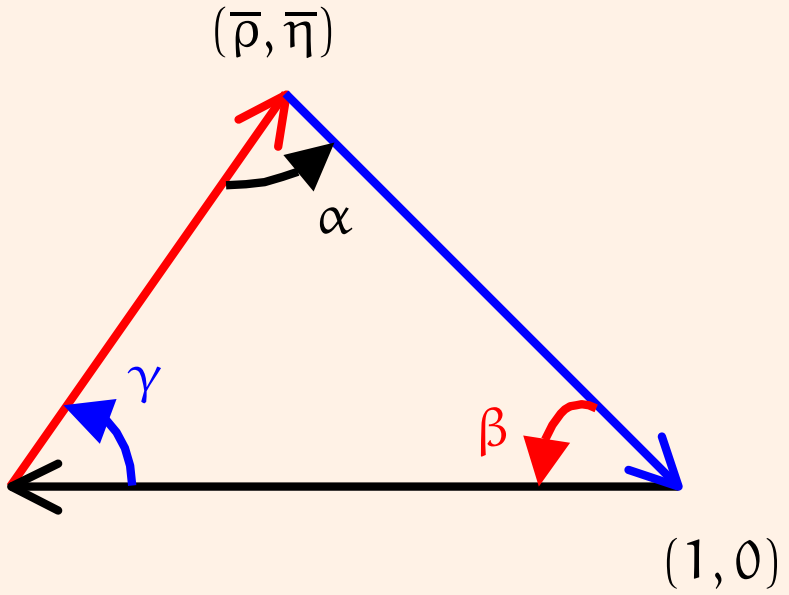
1ère $\leftrightarrow$ (resp. 2ème $\leftrightarrow$ 3rd) génération

$\propto \lambda \sim 0.22$ (resp. $\propto \lambda^2$)

1st $\leftrightarrow$ 3ème génération $\propto \lambda^3$

unitarité CKM $\Rightarrow$ six triangles dans le plan complexe, dont quatre sont quasi plats, et deux sont non plats mais quasi dégénérés



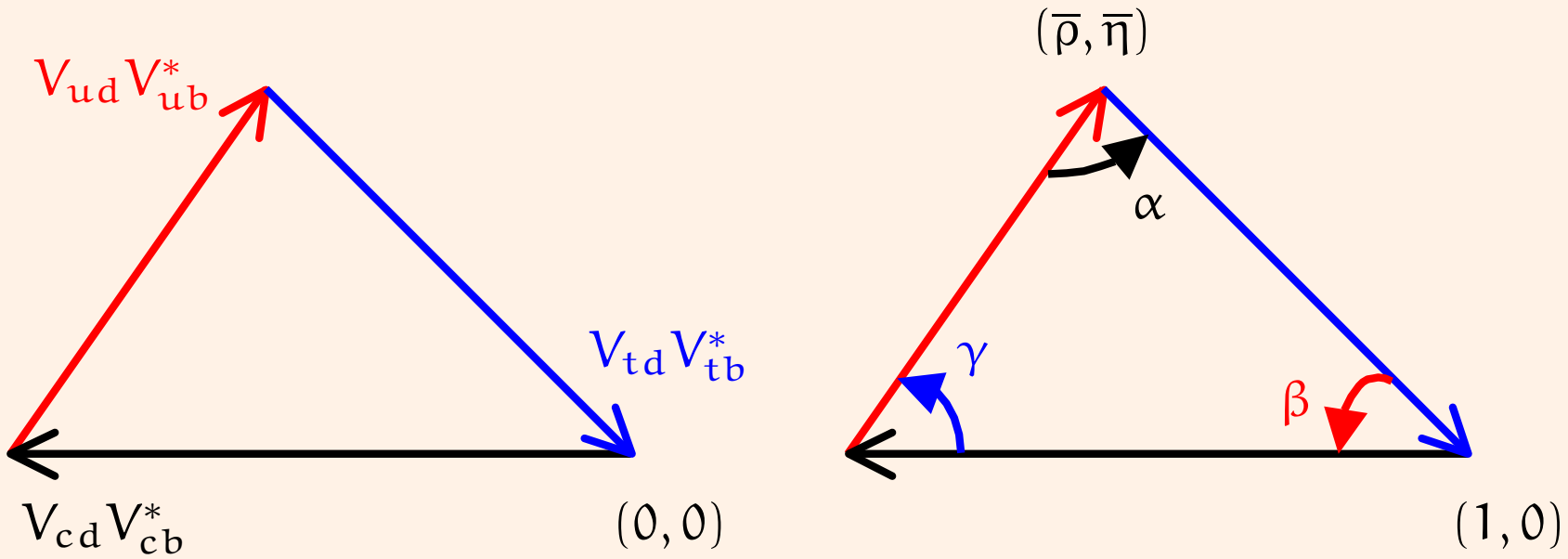


paramétrisation de la matrice CKM à la Wolfenstein, version exacte et indépendante de convention de phase :

$$\lambda^2 \equiv \frac{|V_{us}|^2}{|V_{ud}|^2 + |V_{us}|^2} \qquad A^2\lambda^4 \equiv \frac{|V_{cb}|^2}{|V_{ud}|^2 + |V_{us}|^2}$$

il n'y a pas besoin de s'arrêter à l'ordre $\mathcal{O}(\lambda^4)$

$$\bar{\rho} + i\bar{\eta} \equiv -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}$$



Matrice MNS

depuis qu'on sait les neutrinos massifs (oscillations), on sait également qu'ils peuvent se mélanger comme les quarks

matrice MNS

$$U_{MNS} = U_{23} U_{13} U_{12}$$

$$c_{ij} = \cos\theta_{ij}$$

$$s_{ij} = \sin\theta_{ij}$$

$$U_{12} = \begin{pmatrix} c_{12} & s_{12} e^{-i\delta_{12}} & 0 \\ -s_{12} e^{i\delta_{12}} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$U_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} e^{-i\delta_{23}} \\ 0 & -s_{23} e^{i\delta_{23}} & c_{23} \end{pmatrix}$$

$$U_{13} = \begin{pmatrix} c_{23} & 0 & s_{13} e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{13}} & 0 & c_{13} \end{pmatrix}$$

similaire à CKM, avec deux phases CP-impaires additionnelles, dues aux termes de Majorana

Contraindre les paramètres du Modèle Standard

les transitions chargées, en arbres, donnent d'excellentes contraintes (de 10^{-4} à 10^{-1})

les transitions neutres changeant la saveur (FCNC) deux fois, *via* des diagrammes en boucles, donnent également une information de qualité (10-20%) : $\Delta S = \Delta D = 2$, $\Delta B = \Delta D = 2$ et maintenant $\Delta B = \Delta S = 2$

les transitions FCNC changeant la saveur une fois, $\Delta S = \Delta D = 1$, $\Delta B = \Delta D = 1$ et $\Delta B = \Delta S = 1$, ne donnent pas d'aussi bonnes contraintes à cause des incertitudes expérimentales et/ou hadroniques

en ce qui concerne les neutrinos, les difficultés sont d'ordre expérimental; il n'y a pas de données sur les phases CP. Deux des trois angles de mélanges sont bien déterminés, tandis que θ_{13} fait l'objet de programmes futurs

Contraindre les paramètres du Modèle Standard

les transitions chargées, en arbres, donnent d'excellentes contraintes (de 10^{-4} à 10^{-1})

les transitions neutres changeant la saveur (FCNC) deux fois, *via* des diagrammes en boucles, donnent également une information de qualité (10-20%) : $\Delta S = \Delta D = 2$, $\Delta B = \Delta D = 2$ et maintenant $\Delta B = \Delta S = 2$

les transitions FCNC changeant la saveur une fois, $\Delta S = \Delta D = 1$, $\Delta B = \Delta D = 1$ et $\Delta B = \Delta S = 1$, ne donnent pas d'aussi bonnes contraintes à cause des incertitudes expérimentales et/ou hadroniques

en ce qui concerne les neutrinos, les difficultés sont d'ordre expérimental; il n'y a pas de données sur les phases CP. Deux des trois angles de mélanges sont bien déterminés, tandis que θ_{13} fait l'objet de programmes futurs

QCD et l'extraction des couplages CKM

les couplages conservant CP sont extraits d'observables qui dépendent aussi d'éléments de matrice hadroniques, qui doivent être calculés dans la théorie (*e.g.* par exemple simulation de QCD sur réseau)

typiquement, $\Gamma \sim |V_{CKM}|^2 \, |\langle f \,|\, O \,|\, i \rangle|^2$

en principe les angles CKM CP-impairs peuvent être déterminés de quantités expérimentales uniquement

typiquement,

$$\begin{aligned} A_{CP} &\sim \mathrm{Im}(V_{CKM}^*/V_{CKM})(\overline{M}_{QCD}/M_{QCD}) \\ &\sim \mathrm{Im}(V_{CKM}^*/V_{CKM}) \end{aligned}$$

Exemples

[SM] : probablement dominé par SM

SM

$B \to \pi\pi \quad \alpha$

$B \to DK \quad \gamma$

Exemples

[SM] : probablement dominé par SM

[$\leftarrow$ QCD] : nécessite le calcul d'éléments de matrice hadroniques

SM

$B \to \pi\pi \quad \alpha$

$B \to DK \quad \gamma$

SM $\leftarrow$ QCD

$B(b) \to D(c)\ell\nu \quad |V_{cb}| \leftarrow f_+^{BD}$ (OPE)

$B(b) \to \pi(u)\ell\nu \quad |V_{ub}| \leftarrow f_+^{B\pi}$ (OPE)

$K \to \ell\nu \quad |V_{us}| \leftarrow f_K$

Exemples

[SM] : probablement dominé par SM

[$\leftarrow$ QCD] : nécessite le calcul d'éléments de matrice hadroniques

[NP] : contributions NP potentiellement grandes

SM

$B \to \pi\pi \quad \alpha$

$B \to DK \quad \gamma$

NP

$B \to J/\psi K_s \quad \beta$

$B_s \to J/\psi\phi \quad \beta_s$

$K \to \pi\nu\bar{\nu} \quad \overline{\rho}, \overline{\eta}$

SM $\leftarrow$ QCD

$B(b) \to D(c)\ell\nu \quad |V_{cb}| \leftarrow f_+^{BD}$ (OPE)

$B(b) \to \pi(u)\ell\nu \quad |V_{ub}| \leftarrow f_+^{B\pi}$ (OPE)

$K \to \ell\nu \quad |V_{us}| \leftarrow f_K$

Exemples

[SM] : probablement dominé par SM

[$\leftarrow$ QCD] : nécessite le calcul d'éléments de matrice hadroniques

[NP] : contributions NP potentiellement grandes

SM

$B \to \pi\pi \quad \alpha$

$B \to DK \quad \gamma$

NP

$B \to J/\psi K_s \quad \beta$

$B_s \to J/\psi\phi \quad \beta_s$

$K \to \pi\nu\bar{\nu} \quad \overline{\rho}, \overline{\eta}$

SM $\leftarrow$ QCD

$B(b) \to D(c)\ell\nu \quad |V_{cb}| \leftarrow f_+^{BD}$ (OPE)

$B(b) \to \pi(u)\ell\nu \quad |V_{ub}| \leftarrow f_+^{B\pi}$ (OPE)

$K \to \ell\nu \quad |V_{us}| \leftarrow f_K$

NP $\leftarrow$ QCD

$\varepsilon_K \quad \overline{\rho}, \overline{\eta} \leftarrow B_K$

$\Delta M_{d,s} \quad |V_{tb}V_{td,s}| \leftarrow B_B$

$B \to \ell^+\ell^- \quad |V_{td,s}| \leftarrow f_B$

Spécificités des analyses en physique des saveurs

présence d'erreurs théoriques potentiellement grandes, comparables voire supérieures aux erreurs d'origine purement statistique; il n'y a pas de définition non ambiguë de ces erreurs, ni d'approche systématique pour les traiter $\Rightarrow$ nécessité d'un modèle
(problème probablement négligeable dans le cas des neutrinos)

l'extraction des phases peut en principe être exempte d'incertitudes théoriques; cependant le prix à payer est l'existence de fortes non-linéarités :

- ambiguïtés discrètes (solutions miroir) dues aux fonctions trigonométriques
- présence de frontières physiques, e.g. $|\sin 2\beta| \leq 1$
- fonctions compliquées des paramètres
- biais dus au nombre de degrés de liberté qui peut être mal défini (on ne peut pas mesurer la phase d'une amplitude dont le module est compatible avec zéro)

NB: le problème des ambiguïtés discrètes, qui se pose aussi dans le cas des nombreux paramètres du MSSM, est peu décrit dans la littérature statistique

Le projet CKMfitter, historique, objectifs

Le groupe

JC, théorie, Marseille

Olivier Deschamps, LHCb, Clermont-Ferrand

Sébastien Descotes-Genon, théorie, Orsay

Ryosuke Itoh, Belle, Tsukuba

Andreas Jantsch, ATLAS, Munich

Christian Kaufhold, théorie, Annecy-le-Vieux

Heiko Lacker, ATLAS, Berlin

Stéphane Monteil, LHCb, Clermont-Ferrand

Valentin Niess, LHCb, Clermont-Ferrand

Jose Ocariz, BaBar, Paris

Stéphane T'Jampens, LHCb, Annecy-le-Vieux

Vincent Tisserand, BaBar, Annecy-le-Vieux

Karim Trabelsi, Belle, Tsukuba

http://ckmfitter.in2p3.fr

La préhistoire...

première analyse CKM en 1992 par Schubert et Schmidtler; les erreurs théoriques sont considérées sur le même plan que les erreurs expérimentales (aussi: Ali et al.)

les ateliers préparant à l'expérience BaBar (1996-1997) ont montré le besoin d'une approche plus spécifique (scan method, groupe d'Orsay)

un traitement bayesien est apparu en 1995 (groupes de Rome/Orsay)

le projet CKMfitter a démarré en 2001 (Höcker, Lacker, Laplace, Le Diberder à Orsay)

Scan method

S. Plaszczynski et M.-H. Schune, 1997

idée : si on connaissait parfaitement la partie théorique, on pourrait faire une analyse purement statistique

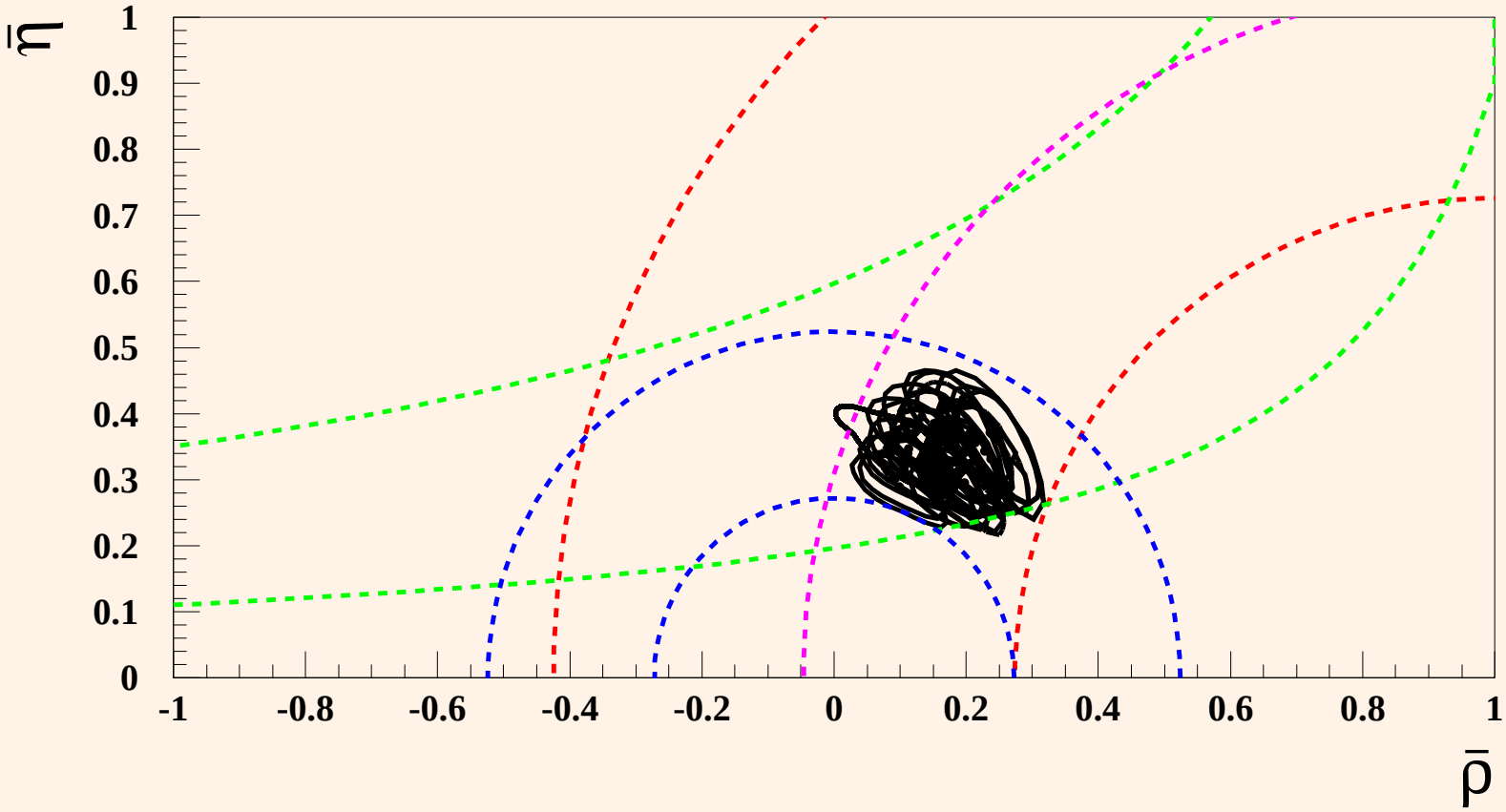
$\Rightarrow$ considérer chaque point possible dans l'espace des paramètres théoriques comme un modèle pour lequel on fait une analyse standard; ne sachant pas quel "modèle" est le bon on scanne uniformément l'espace théorique pour en prendre l'enveloppe

plus précisément, on choisit une valeur de coupure p_c (par exemple $p_c = 0.05$ pour 95% CL, ou $p_c = 0.0027$ pour trois déviations standard), et on ne retient que les "modèles" qui ont un χ^2_{min} correspondant à une p-valeur supérieur à p_c. Pour chaque modèle on calcule la région (hypercontour) correspondant au niveau de confiance choisi, le résultat pour l'ensemble des modèles passant la coupure étant défini comme l'enveloppe des hypercontours

avantage: intuitif et transparent, conservatif

inconvénient: effet de seuil dû au choix de la coupure p_c

Scan method, 1999



Le modèle Rfit

A. Höcker et al., 2001

CKMfitter est né avec l'intention de proposer une hypothèse bien définie concernant les erreurs théoriques, et d'en déduire un traitement fréquentiste rigoureux

Rfit : pour un paramètre $x = x_0 \pm \delta_x$ affecté d'une erreur d'origine purement théorique δ_x, on suppose l'hypothèse $x \in [x_0 - \delta_x, x_0 + \delta_x]$ vérifiée, et rien de plus

avantage: hypothèse simple; conservative à "1σ"

inconvénients: hypothèse formelle, jamais réalisée en pratique; parfois "trop" conservative (grand nombre de paramètres Rfit); risque de sous-estimation de l'erreur à "2σ" et au-delà

empiriquement on constate que cela fonctionne raisonnablement

techniquement, il "suffit" d'écrire $x = x0 + \delta_x \sin\theta_x$ et de considérer θ_x comme un paramètre de nuisance totalement libre

Ajustement de données sur un modèle non linéaire

soit $p = (A, \lambda, \bar{\rho}, \bar{\eta}, \ldots)$ le vecteur des paramètres nécessaires à calculer les observables dans le cadre d'un modèle (ici: SM)

on construit le logarithme de la vraisemblance

$$\chi^2(p) \equiv -2 \ln \mathcal{L}(p) = \sum_{\mathcal{O}} \left(\frac{\mathcal{O}(p) - \mathcal{O}_{exp}}{\sigma_{\mathcal{O}}} \right)^2$$

si les mesures $\mathcal{O}_{exp}$ sont distribuées normalement (très souvent le cas; hypothèse faite dans le reste du cours)

$\chi^2_{min} = \mathrm{Min}_p \chi^2(p)$ peut servir à évaluer la « qualité » du fit; considéré comme une fonction des $\mathcal{O}_{exp}$, χ^2_{min} est asymptotiquement distribué comme un χ^2 avec $N_{dof} = \dim(\mathcal{O}) - \dim(p)$

$\Delta\chi^2(p)$, considéré comme une fonction des $\mathcal{O}_{exp}$, est asymptotiquement distribué comme un χ^2 avec $N_{dof} = \dim(p)$ et peut servir à construire un niveau de confiance (exo: trouver le piège)

correspondance entre CL et $\Delta\chi^2$

	CL	$\dim(p)=1$	$\dim(p)=2$
«1σ»	68.3%	1	2.30
«2σ»	95.5%	4	6.18
«3σ»	99.7%	9	11.8
«5σ»	$1-5.73\times10^{-7}$	25	28.7

en pratique la dimension de p est souvent élevée, et on s'intéresse à un sous-ensemble $p=(\gamma,\mu)$ où γ représente le(s) paramètre(s) d'intérêt, et μ les paramètres de nuisance

on construit alors

$$\Delta\chi^2(\gamma)=\mathrm{Min}_\mu\chi^2(\gamma,\mu)-\chi^2_{\mathrm{min}}$$

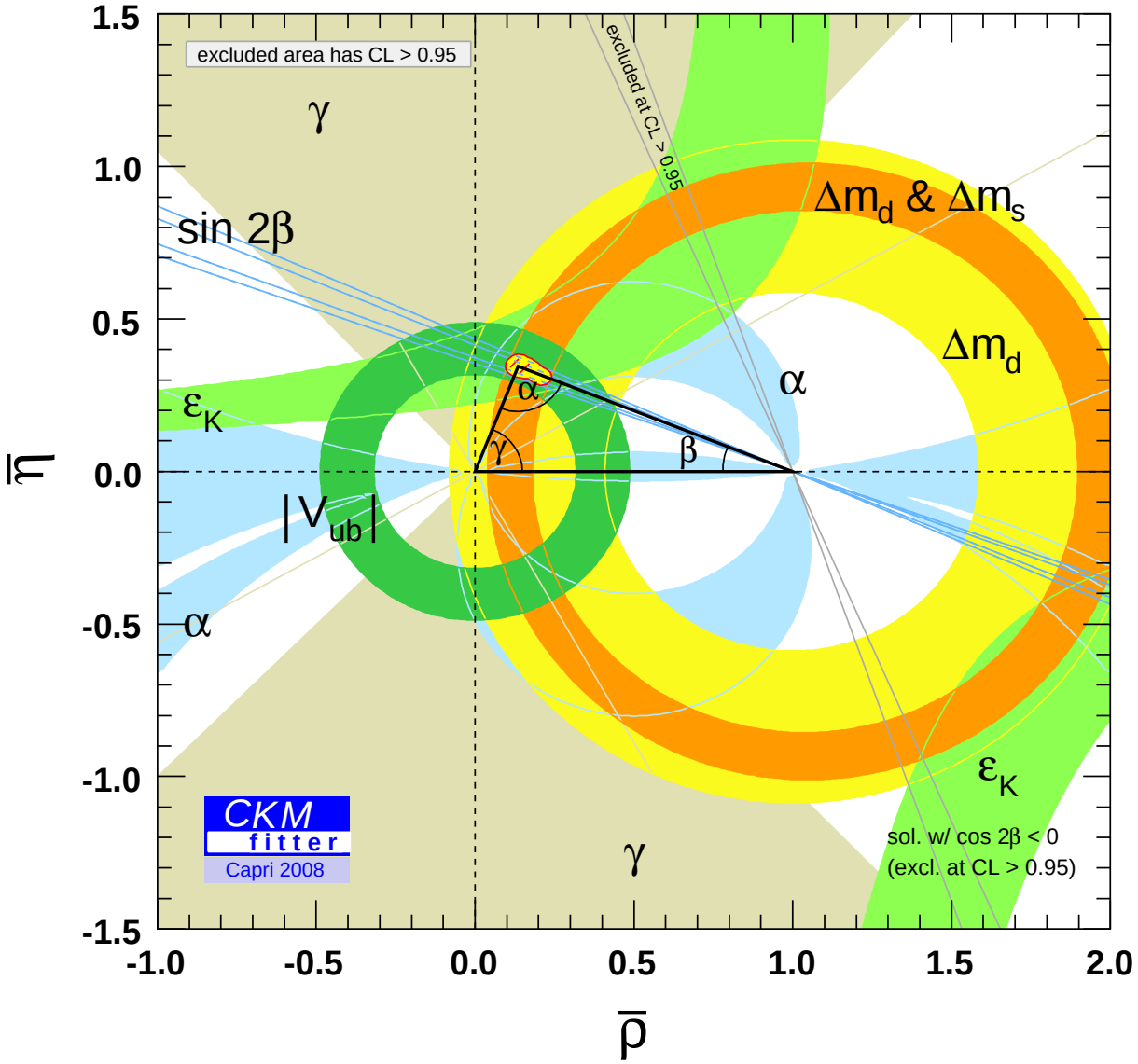
qui est asymptotiquement distribué comme un χ^2 avec $N_{\mathrm{dof}}=\dim(\gamma)$

l'estimation du niveau de confiance sur γ à partir du $\Delta\chi^2(\gamma)$ constitue la méthode par défaut dans CKMfitter

c'est également la procédure suivie par MINOS dans MINUIT

techniquement cela demande, pour chaque valeur de γ, d'effectuer au moins une minimisation (plus la minimisation globale): coût $\sim N_\gamma$

Exemple: l'ajustement CKM global



régions à 95% CL, hiver 2008

11 mesures CKM,

16 paramètres à ajuster (dont 11 mesurés/calculés)

$A = ?$

$\lambda = ?$

$\bar{\rho} = ?$

$\bar{\eta} = ?$

$N_{dof} = \dim(\bar{\rho}, \bar{\eta}) = 2$?

dans l'exemple précédent, chaque contrainte individuelle correspond en fait à $N_{dof} = 1$

$$S_{J/\psi K_S} = \sin 2\beta(\bar{\rho}, \bar{\eta})$$

n'apporte pas d'information séparément sur les deux paramètres

il existe des situations ambiguës

$$\begin{aligned} \mathcal{O}_1 &= f(\bar{\rho}, \bar{\eta}) \\ \mathcal{O}_2 &= g(\bar{\rho}, \bar{\eta}) \end{aligned}$$

apportent *a priori* deux informations indépendantes sur les deux paramètres ... sauf si $g = f + \epsilon$ avec $\epsilon \ll \sigma_{\mathcal{O}_{1,2}}$; dans ce cas N_{dof} ne vaut ni 1 ni 2

de telles situations approximativement dégénérées sont relativement courantes, et parfois indétectables au vu des seules formules analytiques

Le code CKMfitter, présentation et fonctionnalités

le code original, utilisé jusqu'en 2005-2006, est écrit en Fortran et utilise MINUIT pour la minimisation (et était distribué librement)

très beau code dont la modularité a permis l'enrichissement rapide en nouvelles analyses

pour faciliter l'utilisation, ce code est muni d'un « dico » qui permet de faire la conversion entre une quantité physique référencée par une chaîne de caractères, et la variable du code correspondante

le dico rend le code très lent, et de plus en plus lent au fur et à mesure que l'on rajoute de nouvelles quantités physiques

de plus l'utilisation de MINUIT n'est pas optimale, puisqu'il effectue par défaut des tâches dont CKMfitter n'a pas besoin et/ou en a besoin différemment

enfin le langage Fortran n'est pas suffisamment évolué pour permettre facilement la manipulation de structures complexes

Le problème de la minimisation

la minimisation numérique est une procédure complexe, quantité de méthodes existent avec différents avantages et inconvénients; pour les ajustements du type χ^2/vraisemblance sur des modèles non linéaires, ce sont les minimiseurs à métrique variable qui sont les plus utilisés : MIGRAD dans MINUIT

MIGRAD

This is the best minimizer for nearly all functions. It is a variable-metric method with inexact line search, a stable metric updating scheme, and checks for positive-definiteness. It will run faster if you SET STRATEGY 0 and will be more reliable if you SET STRATEGY 2 (although the latter option may not help much). Its main weakness is that it depends heavily on knowledge of the first derivatives, and fails miserably if they are very inaccurate. If first derivatives are a problem, they can be calculated analytically inside FCN (see elsewhere in this writeup) or if this is not feasible, the user can try to improve the accuracy of Minuit's numerical approximation by adjusting values using the SET EPS and/or SET STRATEGY commands (see Floating Point Precision and SET STRATEGY).

(F. James, documentation MINUIT)

idée : puisque les observables dans notre cas sont des quantités calculées analytiquement en fonction des paramètres, pourquoi ne pas également en calculer les dérivées partielles, et en déduire ainsi le gradient exact du χ^2 ou vraisemblance ? on peut espérer ainsi une convergence plus rapide, et surtout une meilleure précision

à la main : trop fatiguant; automatiquement : langage formel, type Mathematica, Maple, Maxima ...

approche alternative : différentier directement le programme lui-même, étape par étape; c'est la différenciation automatique (http://www.autodiff.org), non utilisée dans CKMfitter

autre idée : MINUIT est davantage un logiciel complet qu'une simple routine de minimisation; de nombreuses routines différentes sont utilisées par d'autres communautés $\Rightarrow$ choix d'une routine compacte spécialisée (dmng, D.M. Gay 1980, disponible sur NetLib); algorithme identique à MIGRAD

La structure de CKMfitter

choix de Mathematica pour la partie « physique » et la partie « noyau » qui définit et pilote l'analyse à effectuer

avantages : langage très puissant; très répandu chez les théoriciens

inconvénients : logiciel commercial coûteux; peu utilisé par les expérimentateurs

choix de Fortran pour la partie minimisation et ajustement, approches Monte-Carlo

avantages: langage éprouvé et très rapide, d'excellents compilateurs sont disponibles

inconvénient: langage archaïque

séparation des tâches: ce qui coûte en temps en Fortran, ce qui coûte en complexité en Mathematica

éventuellement, pour les sorties graphiques on utilise ROOT

Mathematica theory packages

observables = F(parameters)
+ gradient

Fortran code

scan & minimization
Monte-Carlo

CKM fitter

Mathematica library

"compactification" algorithm
Fortran routine writer
analysis driver

tools

additional facilities

L'algorithme de « compactification »

n'importe quel langage de calcul formel sait calculer des dérivées partielles analytiquement

on s'aperçoit cependant que la taille des expressions obtenues est souvent très grande : le calcul numérique de la valeur du gradient exact peut finalement être plus coûteux que le calcul approché par différences finies

c'est pour remédier à ce problème que la différentiation automatique a été développée

technique alternative dans CKMfitter : construire automatiquement les sous-expressions communes à la fonction et ses dérivées partielles, et les calculer itérativement en les remplaçant au fur et à mesure dans les sous-expressions suivantes

exemple : on appelle $k_1 = A^2\lambda^4$, calculé une fois pour toutes, et remplacé dans tous les expressions qui le contiennent

L'algorithme de « compactification »

n'importe quel langage de calcul formel sait calculer des dérivées partielles analytiquement

on s'aperçoit cependant que la taille des expressions obtenues est souvent très grande : le calcul numérique de la valeur du gradient exact peut finalement être plus coûteux que le calcul approché par différences finies

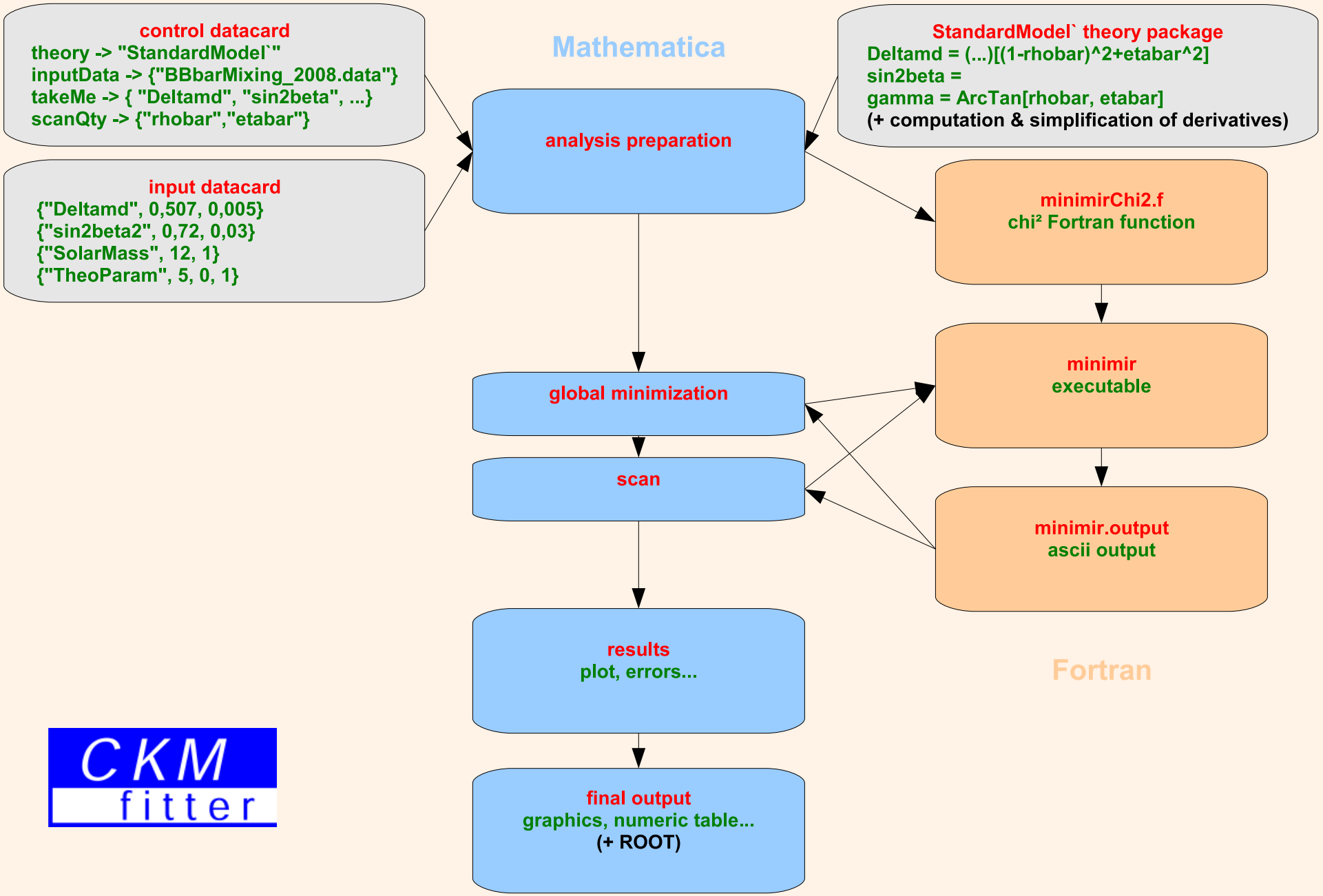
c'est pour remédier à ce problème que la différentiation automatique a été développée

technique alternative dans CKMfitter : construire automatiquement les sous-expressions communes à la fonction et ses dérivées partielles, et les calculer itérativement en les remplaçant au fur et à mesure dans les sous-expressions suivantes

exemple : on appelle $k_1 = A^2\lambda^4$, calculé une fois pour toutes, et remplacé dans tous les expressions qui le contiennent

au final, l'utilisation du gradient exact, de l'algorithme de compactification, et de quelques autres astuces fait gagner à la nouvelle version de CKMfitter un facteur 3–10 par rapport à un code traditionnel dédié (à l'analyse en question), et certainement beaucoup plus par rapport à un code modulaire non dédié ... on a constaté 2, voire 3 ordres de grandeur gagnés par rapport à l'ancienne version Fortran de CKMfitter, ce qui permet de nouvelles analyses quasi-impossibles autrement

Le flot de calcul



Les différents types de données expérimentales/théoriques

article | discussion | edit | history

CKMfitter inputs

CKMfitter can handle 9 types of experimental or theoretical inputs. Here is the list and corresponding syntax.

input type	meaning	syntax
"Fixed"	x = 12	{"x",12}
"Gauss"	x = 12 +-4(stat)	{"x",12,4}
"GaussAsym"	x = 12 +5(stat)-4(stat)	{"x",12,5,-4}
"Range"	x = 12 +-3(theo)	{"x",12,0,3}
"GaussRange"	x = 12 +-4(stat) +-3(theo)	{"x",12,4,3}
"GaussAsymRange"	x = 12 +5(stat)-4(stat) +-3(theo)	{"x",12,5,-4,3}
"Correlation"	correlation matrix (upper triangle)	{ {"x","y","z"}, {1,-0.2,0.1, 1,0.3, 1} }
"UpperLimit"	0 < x < 3.4 10^-5 @ 90% CL	{"x",{3.4 10^-5},90}
"LUT"	LookUp Table for chi^2	{"x","LUTfile.dat"} (1D) or {{"x","y"},"LUTfile.dat"} (2D)

Remarks:

- the correlation matrix is only defined for symmetrical Gaussian quantities
- correlation submatrices can be defined transparently, the full correlation matrix being constructed automatically
- shortcuts/aliases can be defined by {"mySet", "input1", input2",...} so that "mySet" will correspond to the specified list of individual inputs. Shortcuts can currently **not** refer to other shortcuts.
- a LUT file is a chi^2 ASCII file with a specific syntax

Les instructions de contrôle

OVERVIEW

Name	Comment	Optional?
analysisName	analysis reference name	mandatory
theoryPackage	theory model to fit	mandatory
inputData	experimental and theoretical inputs	mandatory
job	control script datacards	optional
jobName	job reference name inside an analysis	optional
jobLoop	shortcut for 1D script datacards	optional
takeMe	actual list of inputs to have in the fit	mandatory
correlationMatrix	whether to take into account correlation matrix(ces)	optional
changeOfVariables	change of variables on theory parameters	optional
scanQty	the quantity(ies) to be scanned	optional
scanMin, scanMax	scan window	optional
changeOfUnits	change of units on scanned quantities	optional
equivalence	equivalent physical content for different quantities	optional
replaceInput	replace the label of a specific input	optional
startRange	typical ranges for fit parameters	optional
globalMinSearches	number of fit tries for global minimization	optional
nbOfFits	number of fit tries at each scan point	optional
nbOfScans	the number of times the scan window is inspected	optional
scanDirection	direction in which 2D scan is performed	optional
granularity	binning of the scan window	optional
superImpose	superimpose direct input point	optional
CLcut	put in white the 1-CL<CLcut 2D region	optional
outputType	output file type(s)	optional
verbose	print unnecessary information	optional
useOneDof	Ndof = 1 in 2D plots	optional
substractChi2Min	substract global minimum chi2	optional
gaussOption	apply Gaussian errors	optional
errorScaleFactor	used in conjunction with gaussOption	optional

Détecter les ambiguïtés discrètes

avec un minimiseur local, on ne peut jamais être certain d'avoir convergé vers le minimum global

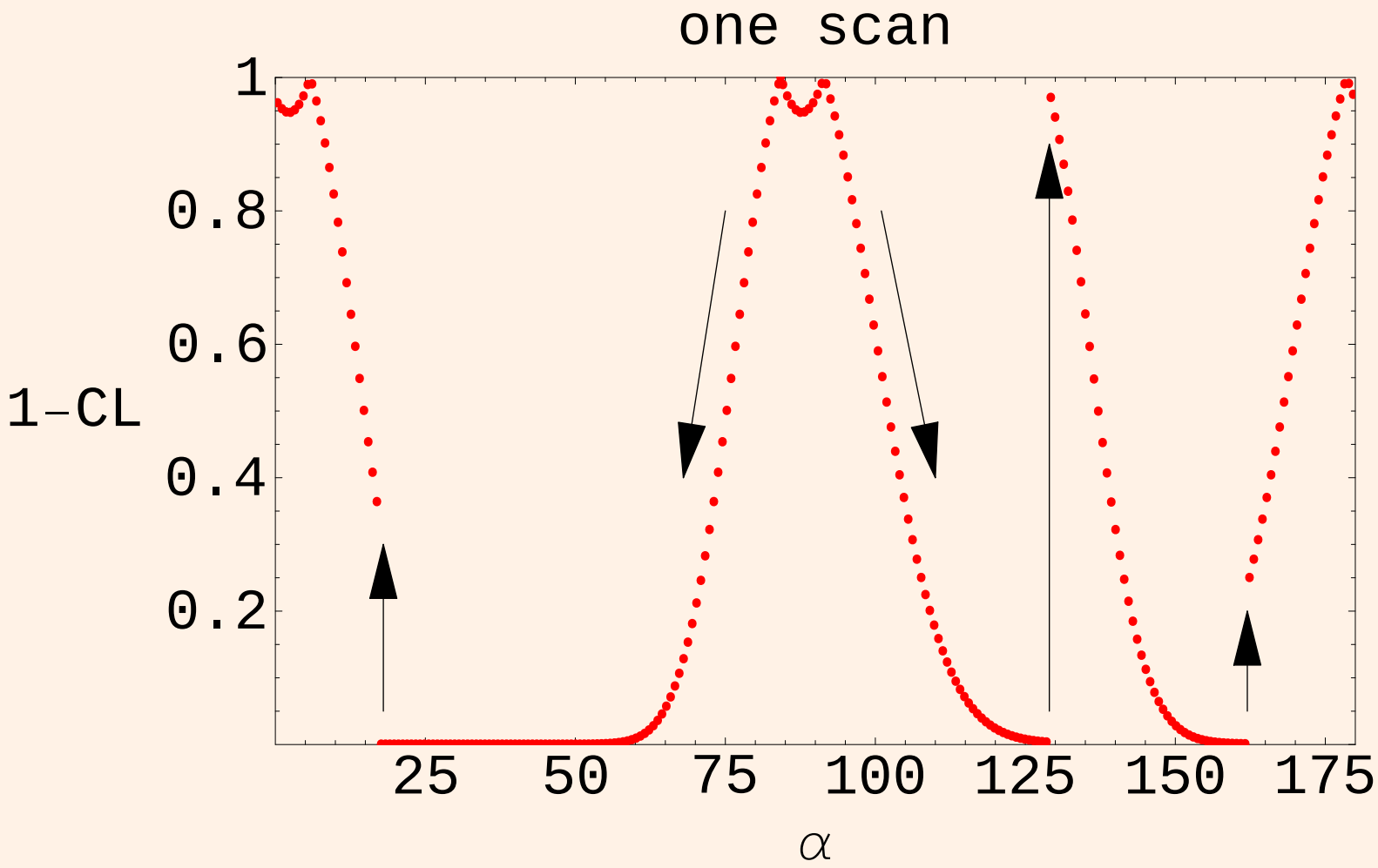
souvent en redémarrant le minimiseur avec des valeurs initiales différentes, on peut converger vers un minimum différent; on garde alors le meilleur de tous ceux trouvés

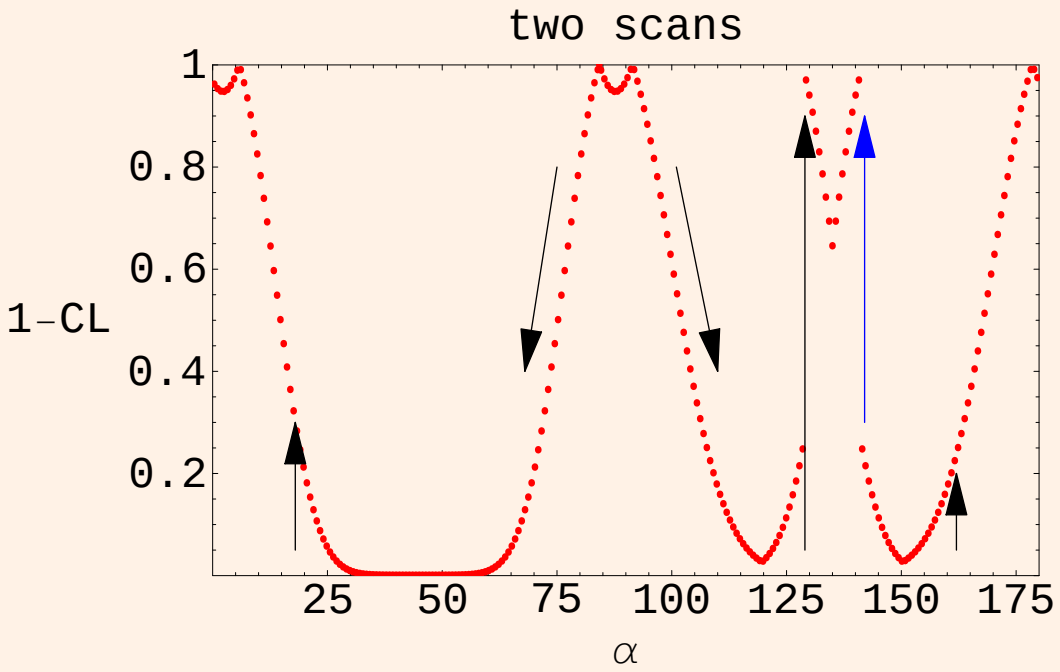
des techniques «non-locales» existent, basées sur des modèles thermodynamiques ou génétiques; non disponible dans CKMfitter

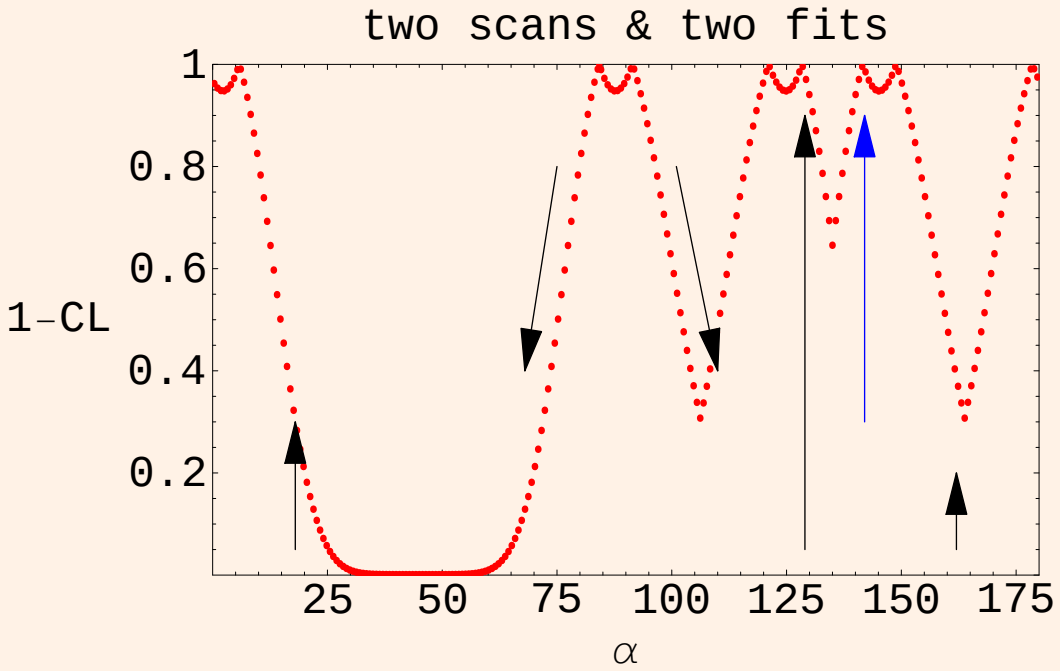
des astuces permettent d'améliorer fortement la convergence

L'extraction de α de $B \to \pi\pi$

on peut montrer géométriquement et analytiquement que l'analyse d'isospin possède en général 8 solutions pour α dans $[0, \pi]$







en scannant plusieurs fois dans plusieurs sens, on parcourt différents minimas locaux

l'idée est toujours de partir des résultats de la minimisation précédente pour initialiser la minimisation suivante

Principales possibilités de CKMfitter

calcul du niveau de confiance en fonction des paramètres d'intérêt, point par point, en 1D et 2D

calcul des erreurs et donc des intervalles de confiance en 1D

représentation par contours ou niveaux de couleurs en 2D

analyse statistique par Monte-Carlo en régime non asymptotique (cf. cours II)

et ce, pour n'importe quel jeu d'observables souhaitées, du moment qu'elles sont codées sous Mathematica :

la matrice CKM dans la paramétrisation Wolfenstein-exacte

le mélange $K\bar{K}$, $B\bar{B}$: ε_K, Δm, $\Delta\Gamma$, A_{SL} (NLO)

$B \to \ell\nu$, $B \to \ell\ell$

$B \to \pi\pi$, $\rho\pi$, $\rho\rho$ dans la limite d'isospin

$B \to \pi\pi$, $K\pi$, $K\bar{K}$ dans la limite SU(3)

$B \to D^{(*)}K^{(*)}$ ADS/GLW/GGSZ

$B \to D^{(*)}\pi(\rho)$

$B \to V\gamma$

$b \to s\gamma$

$K \to \pi\nu\bar{\nu}$ (à mettre à jour)

$B \to \pi\pi$, $K\pi$, $K\bar{K}$ dans la factorisation (à mettre à jour)

Paramètres de nuisance en statistique fréquentiste

Retour sur la notion de couverture

un paramètre p (ou vecteur) est dit appartenir à un intervalle $[p_1, p_2]$ à $\alpha\%$ de niveau de confiance, si la probabilité que l'intervalle $[p_1, p_2]$ contienne la vraie valeur p_{True}, quand on répète un grand nombre de fois l'expérience, vaut α

quand cette probabilité est supérieure à α, on parle d'« overcoverage »; quand elle est inférieure, on parle d' « undercoverage »

la méthode de Neyman permet la construction d'un intervalle de confiance ayant les bonnes propriétes de couverture; le choix de la « fonction d'ordre » est libre, cela peut-être par exemple la PDF(mesures|p) (Neyman), ou encore le rapport de vraisemblance (Feldman-Cousins)

Retour sur la notion de couverture

un paramètre p (ou vecteur) est dit appartenir à un intervalle $[p_1, p_2]$ à $\alpha\%$ de niveau de confiance, si la probabilité que l'intervalle $[p_1, p_2]$ contienne la vraie valeur p_{True}, quand on répète un grand nombre de fois l'expérience, vaut α

quand cette probabilité est supérieure à α, on parle d'« overcoverage »; quand elle est inférieure, on parle d' « undercoverage »

la méthode de Neyman permet la construction d'un intervalle de confiance ayant les bonnes propriétes de couverture; le choix de la « fonction d'ordre » est libre, cela peut-être par exemple la PDF(mesures|p) (Neyman), ou encore le rapport de vraisemblance (Feldman-Cousins)

la construction d'un niveau de confiance peut être formulée dans le cadre du test d'hypothèses, on introduit alors la notion de « p-Valeur »

l'hypothèse à tester s'écrit $H_0 : p_{\text{True}} = p$; on choisit librement un test statistique, dépendant des données et de H_0 (donc de p), tel que plus la valeur du test est petite, plus les données sont en accord avec l'hypothèse

la p-Valeur est alors la probabilité qu'on obtienne une valeur de test moins bonne, si on recommence un grand nombre de fois la même expérience

la p-Valeur étant une fonction des données, est elle-même une variable aléatoire; le critère de couverture est équivalent à demander à ce que la distribution des p-Valeurs, sous l'hypothèse H_0, est uniforme

pourquoi préférer overcoverage à undercoverage ? Supposons qu'on ait $N \gg 1$ tests indépendants du Modèle Standard, du type « mesure = prédiction ». Pour chaque test on construit une p-Valeur, et si le Modèle Standard est vrai, la distribution de ces p-Valeurs doit être uniforme

en pratique on a souvent des hypothèses composites, qui nécessitent la connaissance de paramètres additionnels (paramètres de nuisance) pour pouvoir tout calculer; dans ce cas il n'existe pas de méthode garantissant une couverture exacte

si la distribution des p-Valeurs n'est pas uniforme, et qu'il y a « trop de grandes p-Valeurs », c'est nécessairement que la méthode utilisée «overcovers»

si la distribution des p-Valeurs n'est pas uniforme, et qu'il y a « trop de petites p-Valeurs », c'est soit que l'hypothèse H_0 est fausse (Nouvelle Physique), soit que la méthode utilisée «undercovers»

pour ne pas tirer de conclusions hâtives il est important de préférer overcoverage à undercoverage, dans la limite du raisonnable

Demortier, *p-Values, what they are and how to use them*, CDF note 8622, http://www-cdf.fnal.gov/~luc/statistics/cdf8662.pdf

Formule maître

$$\mathrm{CL}(p) = \int_0^{t(p;\mathrm{data})} \mathrm{d}t\, \mathrm{PDF}_t(t|p)$$

où $t = t(p;\mathrm{data})$ est le test statistique et sa PDF sous H_0 est donnée par

$$\mathrm{PDF}_t(t|p) = \int \mathrm{d}\mathrm{exp}\, \delta[t - t(p;\mathrm{exp})]\, \mathrm{PDF}(\mathrm{exp}|p)$$

et où la PDF des mesures exp en fonction des paramètres p est supposée connue (dans le cours: multigaussienne)

exo: montrer que la distribution des $\mathrm{CL}(p)$ est uniforme, quel que soit p

le choix $t(p;\mathrm{data}) = \mathrm{PDF}(\mathrm{data}|p)$ correspond à la construction de Neyman originale, alors que le choix $t(p;\mathrm{data}) = \Delta\chi^2(p)$ correspond à Feldman-Cousins

si $t = \Delta\chi^2$, asymptotiquement PDF_t est une distribution de χ^2, ne dépend que de $\dim(p)$, et l'intégrale se fait analytiquement (fonctions Γ incomplètes, fonction d'erreur en 1D)

en régime non asymptotique, la forme de PDF_t dépend *a priori* de p, et l'intégrale se fait par méthode Monte-Carlo : nombre de minimisations $\sim N_p \times N_{\mathrm{exp}}$ avec $N_{\mathrm{exp}} \sim 2500$ pour une erreur de ~ 0.01 sur CL

Projection sur un sous-espace

$p = (\gamma, \mu)$, on s'intéresse à γ. La formule maître devient

$$CL_\mu(\gamma) = \int_0^{t(\gamma;\text{data})} dt\, PDF_t(t|\gamma, \mu)$$

l'hypothèse $H_0 : \gamma_{\text{True}} = \gamma$ est en fait composite, elle ne suffit pas à calculer PDF_t, il faut aussi fixer une valeur de μ. La couverture exacte n'est garantie que pour la valeur de μ fixée

si $t = \Delta\chi^2(\gamma)$, asymptotiquement PDF_t est une distribution de χ^2, ne dépend que de $\dim(\gamma)$ et non de γ ni de μ

Projection sur un sous-espace

$p = (\gamma, \mu)$, on s'intéresse à γ. La formule maître devient

$$CL_\mu(\gamma) = \int_0^{t(\gamma;data)} dt\, PDF_t(t|\gamma, \mu)$$

l'hypothèse $H_0 : \gamma_{True} = \gamma$ est en fait composite, elle ne suffit pas à calculer PDF_t, il faut aussi fixer une valeur de μ. La couverture exacte n'est garantie que pour la valeur de μ fixée

si $t = \Delta\chi^2(\gamma)$, asymptotiquement PDF_t est une distribution de χ^2, ne dépend que de $dim(\gamma)$ et non de γ ni de μ

Supremum method

$$CL(\gamma) = Min_\mu CL_\mu(\gamma)$$

garantit la couverture pour toute valeur de μ. Il peut donc y avoir overcoverage pour des valeurs spécifiques de μ

nombre de minimisations du $\Delta\chi^2 \sim 2 \times N_\gamma \times N_{exp} \times N_{MinCL}$ où N_{MinCL} est le nombre d'appels à $CL_\mu(\gamma)$ pour converger vers le minimum

indépendamment de sa complexité technique, la méthode du supremum présente le gros inconvénient de minimiser $CL_\mu(\gamma)$ par rapport à toutes les valeurs de μ, y compris celles en désaccord avec les données

indépendamment de sa complexité technique, la méthode du supremum présente le gros inconvénient de minimiser $\mathrm{CL}_\mu(\gamma)$ par rapport à toutes les valeurs de μ, y compris celles en désaccord avec les données

Plugin method

au lieu de minimiser par rapport à μ, on utilise un estimateur, par exemple le maximum de vraisemblance à γ fixe, $\mu = \hat{\mu}(\gamma)$. Technique beaucoup moins coûteuse, mais qui ne prend pas en compte l'incertitude sur μ et ne garantit pas la couverture

indépendamment de sa complexité technique, la méthode du supremum présente le gros inconvénient de minimiser $\mathrm{CL}_\mu(\gamma)$ par rapport à toutes les valeurs de μ, y compris celles en désaccord avec les données

Plugin method

au lieu de minimiser par rapport à μ, on utilise un estimateur, par exemple le maximum de vraisemblance à γ fixe, $\mu = \hat{\mu}(\gamma)$. Technique beaucoup moins coûteuse, mais qui ne prend pas en compte l'incertitude sur μ et ne garantit pas la couverture

Confidence interval supremum method

on construit d'abord un intervalle à $(1-\beta)\%$ CL pour μ, et on minimise par rapport à μ dans cet intervalle. Alors

$$\mathrm{CL}(\gamma) = \mathrm{Min}_{\mu \in C_\beta} \mathrm{CL}_\mu(\gamma) + \beta$$

est une p-Valeur «valide», qui garantit la couverture (ou overcoverage) (Boos & Berger 1994)

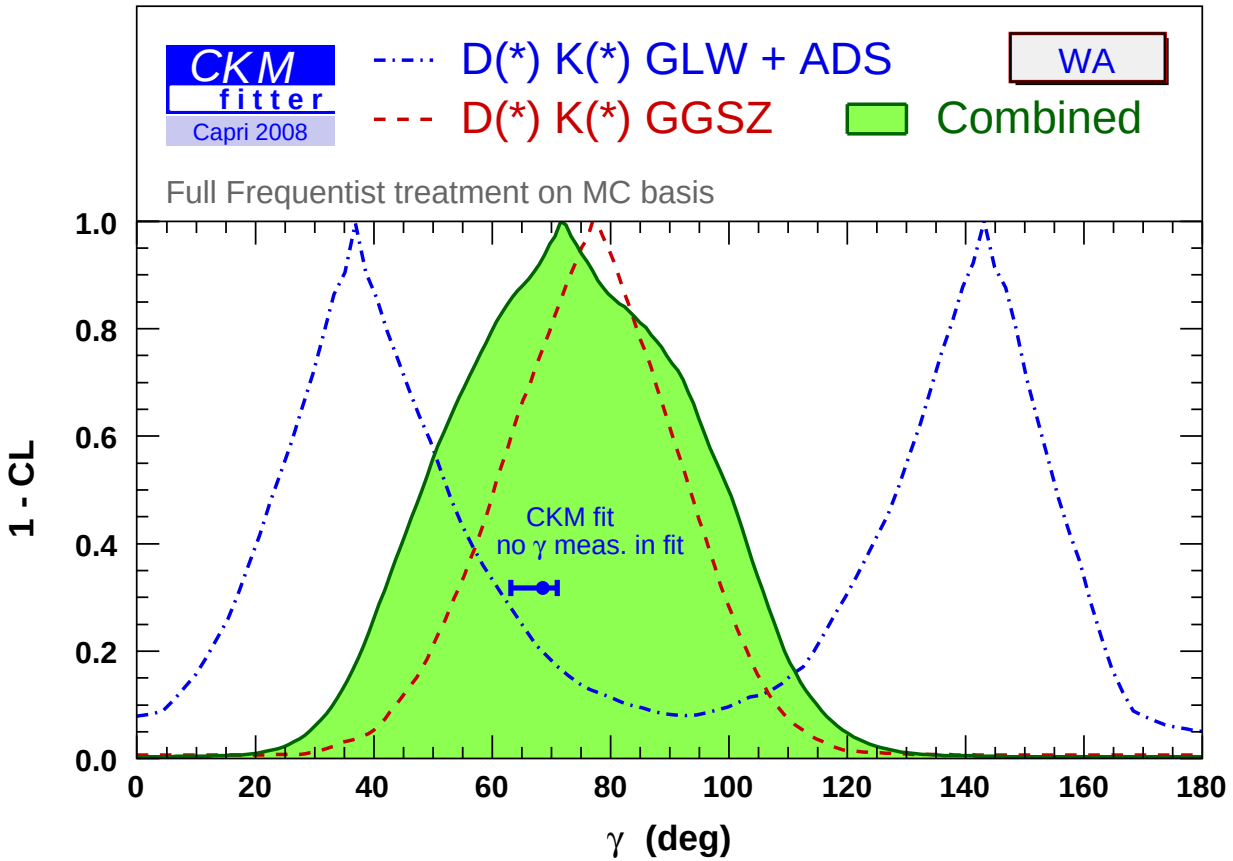
NB: il faut prendre β petit mais pas trop, et le choisir avant de regarder les données

quand la dimension de μ est grande (notre cas ...), c'est techniquement très difficile à implémenter (pas de solution dans CKMfitter pour le moment). De plus, il n'est pas clair que ce soit une amélioration significative par rapport à la méthode du supremum

il existe d'autres méthodes, dont certaines où même le test statistique est évalué par Monte-Carlo... Cependant il ne semble pas y avoir d'approche miracle qui donne des résultats satisfaisants dans tous les cas

il existe d'autres méthodes, dont certaines où même le test statistique est évalué par Monte-Carlo... Cependant il ne semble pas y avoir d'approche miracle qui donne des résultats satisfaisants dans tous les cas

un exemple concret: l'extraction de l'angle γ à partir des désintégrations $B \to DK$. Il s'agit de mesurer l'interférence entre deux amplitudes dont le rapport est petit $r_B \sim 0.10$



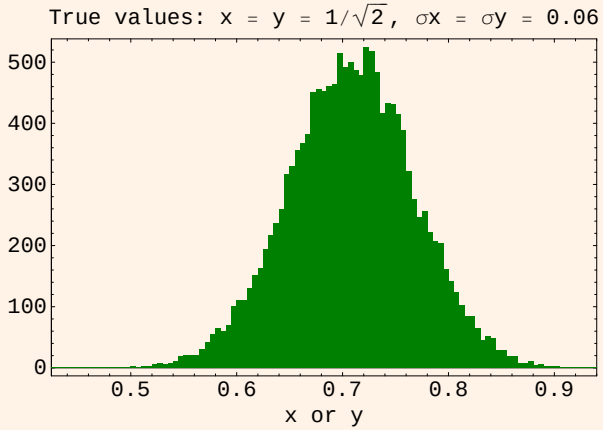
30 mesures, 12 paramètres,
200 points, 2500 toy exp., 4 scans,
12 fits ...

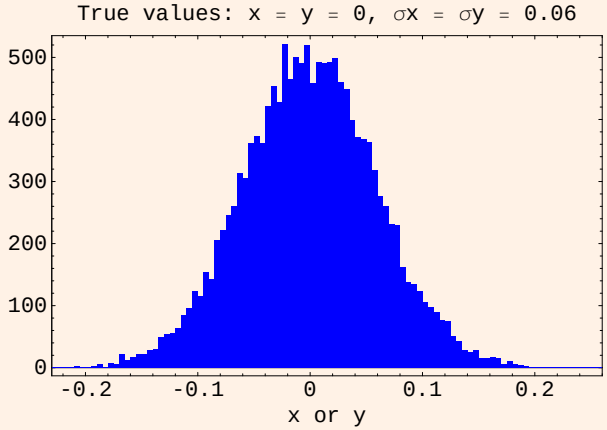
supremum (courbe verte)
$\gamma = (72^{+34}_{-30})^\circ$

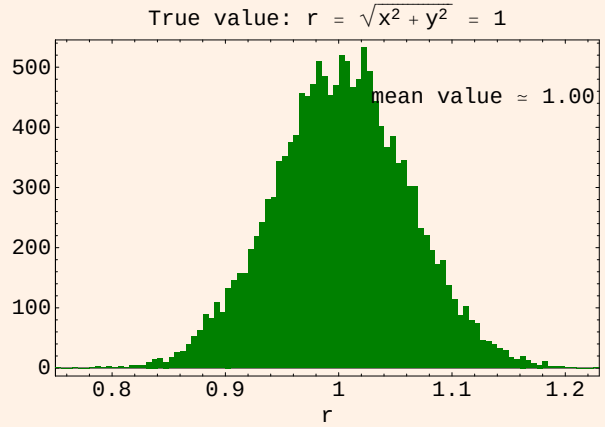
χ^2-distrib
$\gamma = (72^{+18}_{-17})^\circ$

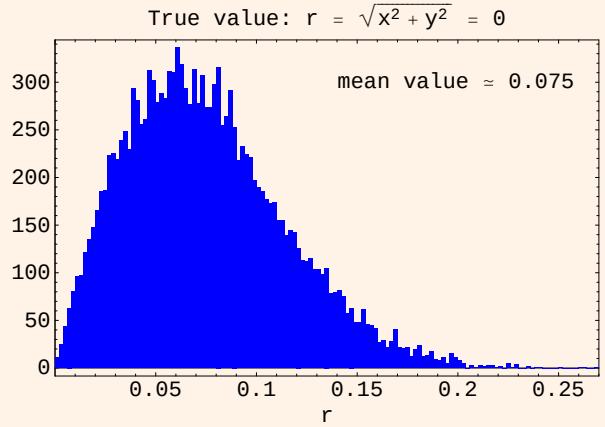
plugin
$\gamma = (72^{+24}_{-18})^\circ$

Biais sur r_B









en haut, distribution de x ou y, normalement distribuée autour de 1 (vert) et 0 (bleu), variance = 0.06

en bas, distribution correspondante de $r = \sqrt{x^2 + y^2}$

avec $x = r\cos\phi$, $y = r\sin\phi$ le maximum de vraisemblance pour r est biaisé si les valeurs vraies de x, y sont proches de zéro

dans l'analyse de l'angle γ, comme l'erreur naïve sur γ est inversement proportionnelle à $1/r_B$, elle est biaisée (sous-estimée) si r_B est biaisé vers les grandes valeurs

on peut aussi voir les choses autrement en remarquant que quand x et y tendent vers zéro, la phase ϕ devient indéterminée et ne contribue plus comme degré de liberté

on peut aussi voir les choses autrement en remarquant que quand x et y tendent vers zéro, la phase ϕ devient indéterminée et ne contribue plus comme degré de liberté

attention, ces arguments sont naïfs : on peut montrer que dans le cas cartésien/polaire à 2D, l'erreur sur ϕ déterminée de $\Delta\chi^2(\phi)$ n'est pas biaisée, bien que l'estimateur de r le soit (à faire en exercice...)

Un exemple partiellement soluble : l'analyse GGSZ

le Modèle Standard prédit, pour chacun des modes $B \to DK, DK^*, D^*K$

$$\begin{aligned} x_\pm &= r_B \cos(\delta \pm \gamma) \\ y_\pm &= r_B \sin(\delta \pm \gamma) \end{aligned}$$

ici les paramètres de nuisance sont $\mu = (r_B, \delta)$

les mesures dans les coordonnées cartésiennes $x_\pm$, $y_\pm$ sont gaussiennes avec une bonne approximation. Si on fait l'hypothèse que toutes les erreurs sont égales $\sigma_{x_\pm} = \sigma_{y_\pm} = \sigma$ (ce qui est assez proche des données réelles) on peut calculer analytiquement $\Delta\chi^2$

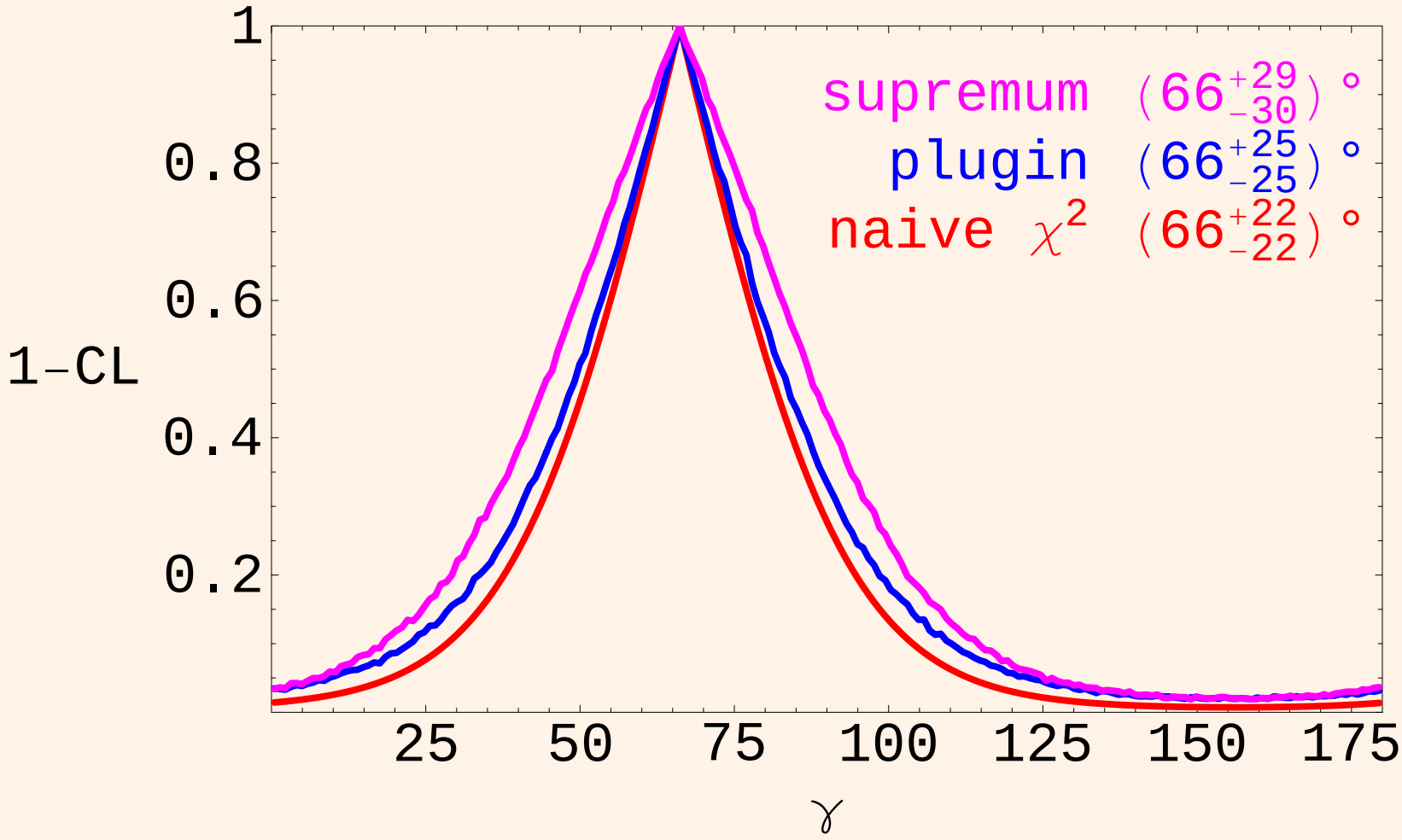
$$\begin{aligned} \Delta\chi^2(\gamma) = \frac{1}{\sigma^2} \Big[& \sqrt{(X_+^2 + Y_+^2) + (X_-^2 + Y_-^2)} \\ & -(X_+X_- + Y_+Y_-)\cos 2\gamma + (X_+Y_- - X_-Y_+)\sin 2\gamma \Big] \end{aligned}$$

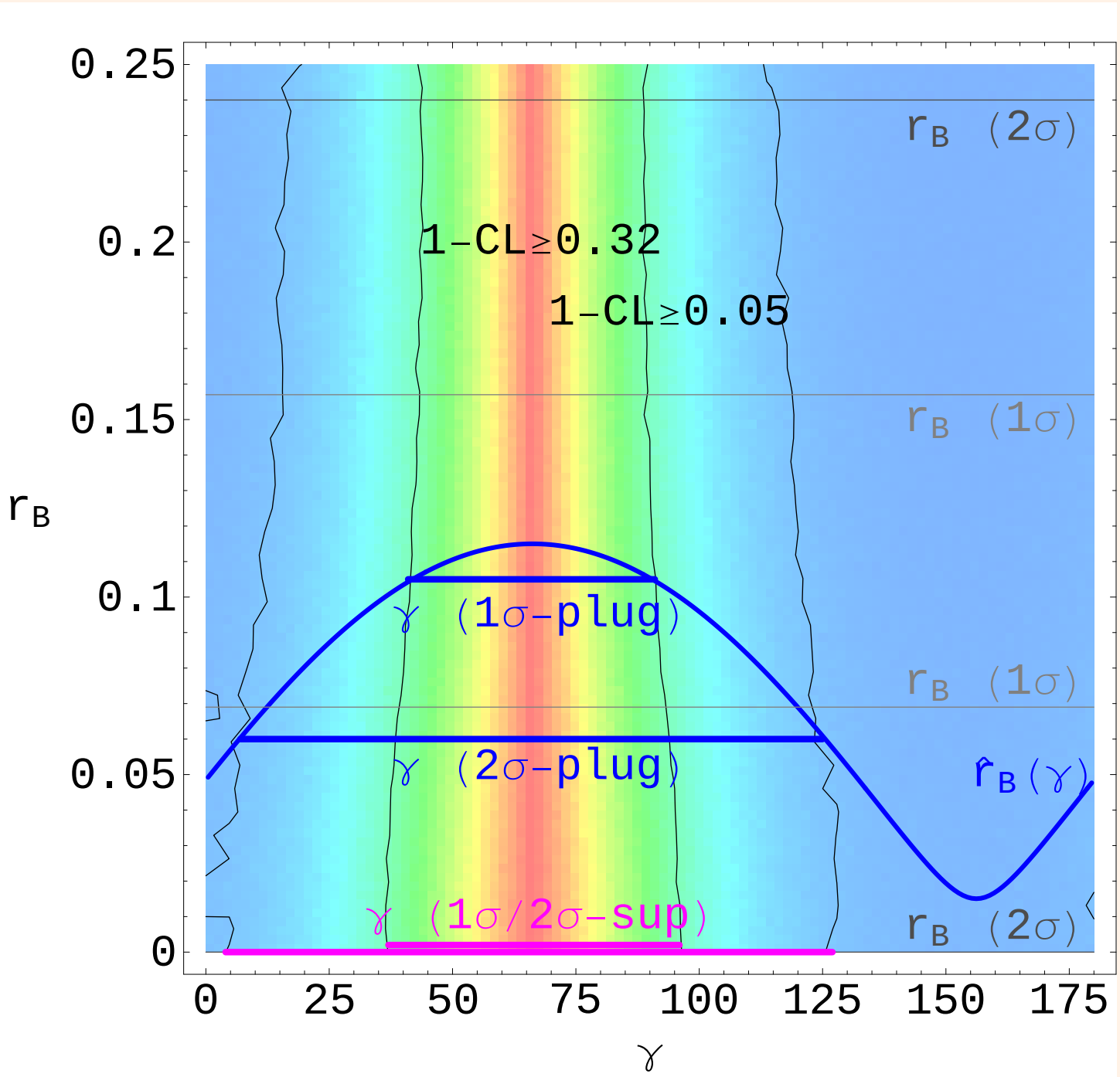
où les $X_\pm$, $Y_\pm$ sont les mesures (réelles ou simulées).

le calcul de l'intégrale dans la formule maître n'est pas possible analytiquement (?), cependant la forme explicite de $\Delta\chi^2$ fait gagner au moins deux minimisations par évaluation du test statistique et permet une étude poussée

Données «réalistes» en 2008

x_+	y_+	x_-	y_-
-0.1287 ± 0.06	0.0186 ± 0.06	0.0772 ± 0.06	0.0635 ± 0.06





Dépendance par rapport à r_B

on construit $CL_{r_B,\delta}(\gamma)$ puis $CL_{r_B}(\gamma) = \text{Min}_\delta CL_{r_B,\delta}(\gamma)$

la dépendance par rapport à r_B est assez douce, il semble artificiel d'imposer $r_B = \hat{r}_B(\gamma)$

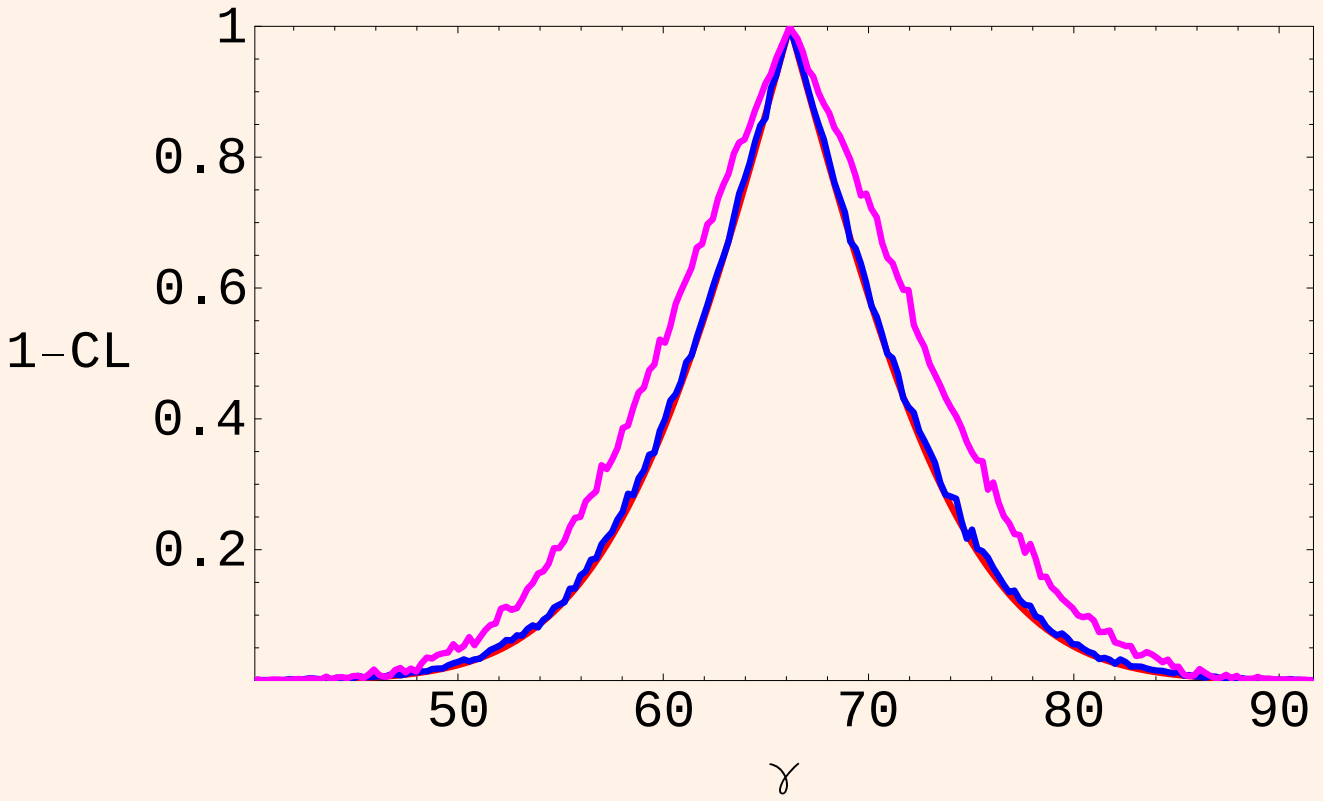
la méthode de Boos & Berger n'est pas très utile, parce que r_B est ici compatible avec zéro à mieux que 95% CL

la valeur supremum de r_B est $r_B = 0$: preuve ?

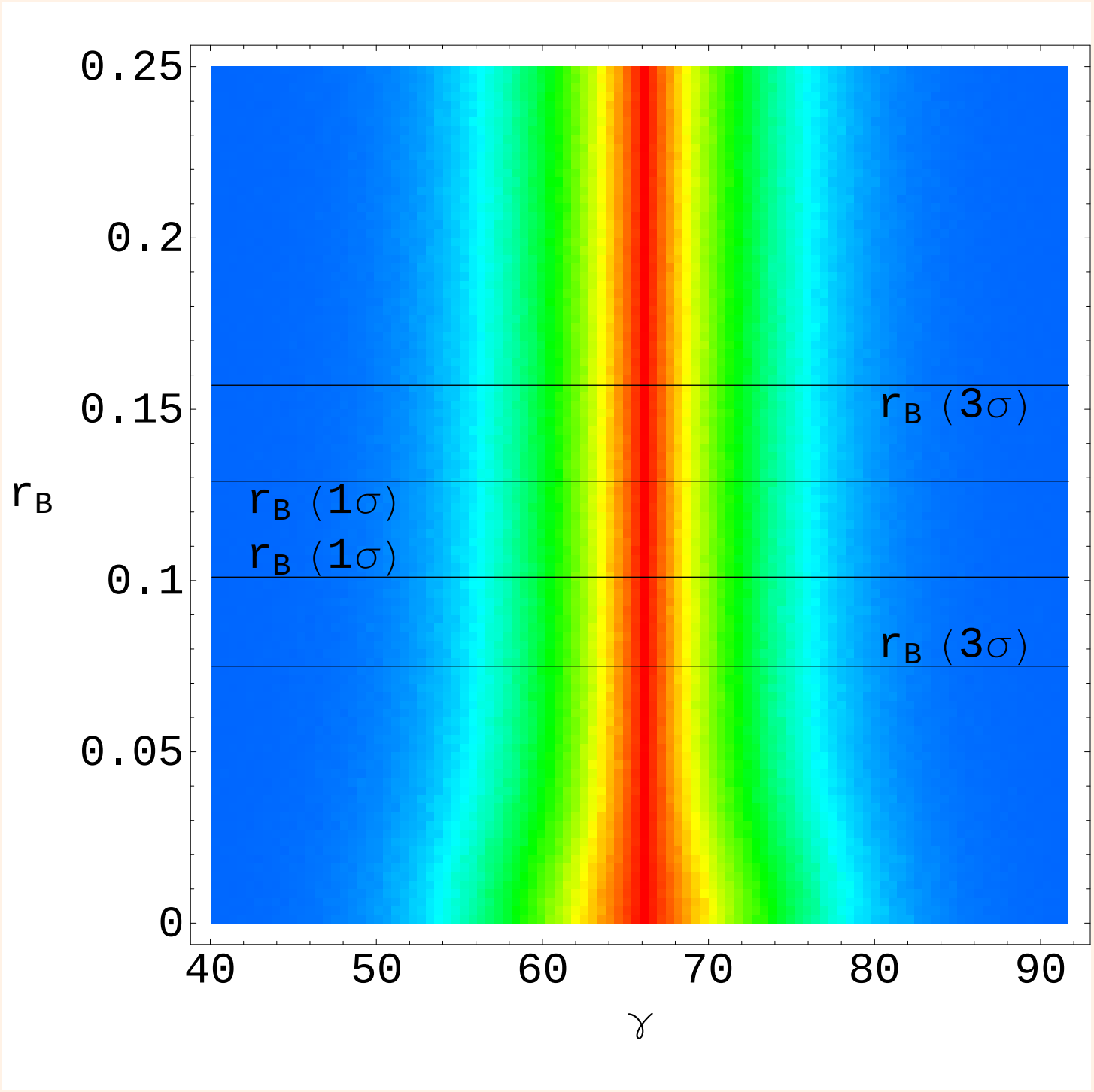
Régime asymptotique ?

on divise les erreurs par 3

x_+	y_+	x_-	y_-
-0.1138 ± 0.02	0.0164 ± 0.02	0.0888 ± 0.02	0.0730 ± 0.02



de manière surprenante, il subsiste une différence entre plugin et supremum...



ici la dépendance par rapport à r_B est beaucoup plus brutale: presque absente pour $r_B > 0.05$, elle devient approximativement linéaire au-delà

en forçant r_B dans son intervalle à 3σ, la distribution du test statistique est un χ^2 à une très bonne approximation

dans cet exemple la méthode supremum apparaît bien trop conservative

Test de couverture

on veut tester l'uniformité de la distribution de la p-Valeur (CL)

Procédure

- on fixe les vraies valeurs γ_{True} et μ_{True}
- on génère un grand nombre d'expériences similaires $X_{\pm}$, $Y_{\pm}$ autour des valeurs vraies
- pour chaque expérience, on calcule CL au point $\gamma = \gamma_{\text{True}}$
- la fraction d'expériences pour lesquelles 1-CL(γ) est supérieur à une valeur α donnée représente la couverture

NB: la couverture dépend *a priori* de γ_{True} et μ_{True}, ainsi que bien sûr de α

Test de couverture

on veut tester l'uniformité de la distribution de la p-Valeur (CL)

Procédure

- on fixe les vraies valeurs γ_{True} et μ_{True}
- on génère un grand nombre d'expériences similaires $X_{\pm}$, $Y_{\pm}$ autour des valeurs vraies
- pour chaque expérience, on calcule CL au point $\gamma = \gamma_{\text{True}}$
- la fraction d'expériences pour lesquelles 1-CL(γ) est supérieur à une valeur α donnée représente la couverture

NB: la couverture dépend *a priori* de γ_{True} et μ_{True}, ainsi que bien sûr de α

exemple : $\alpha = 0.68$

on trouve typiquement que la couverture dans la méthode plugin est de l'ordre de 0.67 (undercoverage), tandis que celle de la méthode supremum est de l'ordre de 0.71 (overcoverage)

résultat peu dépendant des vraies valeurs des paramètres (préliminaire - à confirmer)

L'analyse de γ complète

semble converger plus vite que GGSZ seul : en divisant les erreurs par 3 il y a accord entre les différentes méthodes

le test de couverture de la méthode plugin doit être faisable sur une seule machine puissante

en ce qui concerne le test de couverture de la méthode supremum, il faut probablement un cluster

l'étude de méthodes alternatives est aussi en cours

Démonstration du code

Conclusion et persectives

Physique, ce qui manque dans CKMfitter

$b \to s\ell\ell$

$B \to K^{(*)}\ell\ell$

de nouveaux scénarios de Nouvelle Physique

physique du D

extension à d'autres domaines ?

Conclusion et persectives

Physique, ce qui manque dans CKMfitter

$b \to s\ell\ell$

$B \to K^{(*)}\ell\ell$

de nouveaux scénarios de Nouvelle Physique

physique du D

extension à d'autres domaines ?

Technique et outils statistiques

le code est très optimisé, et permet ainsi d'aborder un grand nombre de problèmes

beaucoup d'astuces algorithmiques peuvent *a priori* être utiles ailleurs

une version publique du code est souhaitable

en ce qui concerne les méthodes statistiques Monte-Carlo, il reste à finaliser les tests de couverture, à reconsidérer l'extraction de l'angle α, à traiter l'ajustement global en présence d'erreurs théoriques

également continuer l'étude des différentes méthodes de traitement des paramètres de nuisance