Let us finish this purely formal introduction of the Lie series with a couple of examples that will be very useful during the discussion of the proof of the KAM theorem. Let $(\underline{p},\underline{q})$ be action-angle variables.

- Translation of the actions

Let $\chi(\underline{q}) = \underline{\xi}\cdot\underline{q}$ being $\underline{\xi}$ a constant vector in $\mathbb{R}^n$

$\forall j=1,\dots,n$:

$$\exp\left(\varepsilon L_{\underline{\xi}\cdot\underline{q}}\right) q_j = q_j \quad \left(\text{because } \{q_j, \underline{\xi}\cdot\underline{q}\} = 0\right)$$

$$\exp\left(\varepsilon L_{\underline{\xi}\cdot\underline{q}}\right) p_j = p_j - \varepsilon\xi_j \quad \left(\text{because } \{p_j, \xi_j q_j\} = -\xi_j\right)$$

- Deformation of the actions

$$\chi(\underline{q}) = X(\underline{q})$$

$\forall j=1,\dots,n$:

$$\exp\left(\varepsilon L_{X(\underline{q})}\right) q_j = q_j \quad \left(\text{because } \{q_j, X(\underline{q})\} = 0\right)$$

$$\exp\left(\varepsilon L_{X(\underline{q})}\right) p_j = p_j - \varepsilon\frac{\partial X}{\partial q_j} \quad \left(\text{because } \{p_j, X(\underline{q})\} = -\frac{\partial X}{\partial q_j}\right)$$

Let us remark that both these generating functions leave unchanged the angles, because they do not depend on the ~~angles~~ actions.

• Formal algorithm constructing the Kolmogorov normal form by using the convergence scheme of the classical series

In order to fix the ideas let us consider a Hamiltonian such that its integrable part is made by ~~an~~ a polynomial in the action that contains just a linear part and a quadratic one; moreover, its perturbative part depends on the angles only and its ~~series~~ Fourier expansion is finite.

Let us write such a Hamiltonian with the following general expression:

number of steps of the algorithm — order in ε (trig. degree) — degree in action

$$H^{(0)} = \underline{\omega}\cdot p + f_2^{(0,0)}(p) + \sum_{l=0}^{2}\sum_{s\geq 1} \varepsilon^{s} f_l^{(0,s)}(p,q)$$

where $f_l^{(0,s)} = 0$ $\forall\, l=1,2$ and $s\geq 1$ or $\forall\, s\geq 2$ when $l=0$, $f_2^{(0,0)}\in\mathcal{P}_{2,0}$ and $f_0^{(0,1)}\in\mathcal{P}_{0,K}$.

Here, we make constantly use of the classes of function $\mathcal{P}_{\cdot,\cdot}$.

<u>Definition</u>: We say that a function $f\in\mathcal{P}_{l,sK}$ if its Taylor-Fourier expansion is such that

$$f(p,q) = \sum_{|\underline{j}|=l}\ \sum_{|k|\leq sK} p^{\underline{j}} \exp(i\underline{k}\cdot q).$$

In order to design the constructive algorithm, the following property of the classes of functions is fundamental.

Lemma: let $f \in \mathcal{P}_{\ell,sK}$ and $g \in \mathcal{P}_{\ell',s'K}$ then

$$\{f,g\} \in \mathcal{P}_{\ell+\ell'-1,(s+s')K}.$$

The proof can be seen as a simple exercise.

It is convenient to represent the expansions with a two-dimensional set of cells

	$\|p\|^2$	$f_2^{(0,0)}$	0	0	...
$H^{(0)} =$	$\|p\|$	$\underline{\omega}\cdot \underline{p}$	0	0	...
	1	0	$\varepsilon f_0^{(0,1)}$	0	...
		$\varepsilon^0 = 1$	$O(\varepsilon)$	$O(\varepsilon^2)$	...

Let us try to remove the main (and only in this basic example) perturbing term, by a deformation of the actions, i.e.

$$\hat{H}^{(1)} = \exp\left(\mathcal{L}_{X^{(1)}}\right) H^{(0)}, \quad \text{with } X^{(1)}(\underline{q}) \in \mathcal{P}_{0,K} \text{ such that}$$

$$\varepsilon\left\{\underline{\omega}\cdot\underline{p}, X^{(1)}\right\} + \varepsilon f_0^{(0,1)} = 0, \quad \text{where } f_0^{(0,1)}(\underline{q}) = \sum_{0<|\underline{k}|\le K} c_{\underline{k}} \exp(i\underline{k}\cdot\underline{q})$$

Therefore, $X^{(1)}(\underline{q}) = \sum_{0<|\underline{k}|\le K} x_{\underline{k}} \exp(i\underline{k}\cdot\underline{q})$ with $x_{\underline{k}}$ such that

$$\left(-i\underline{\omega}\cdot\underline{k}\, x_{\underline{k}} + c_{\underline{k}}\right)\exp i\underline{k}\cdot\underline{q} = 0 \Longrightarrow x_{\underline{k}} = \frac{c_{\underline{k}}}{i(\underline{k}\cdot\underline{\omega})}$$

"small divisor"

Of course, by the application of the Lie series generates two new terms: $\varepsilon\{f_2^{(0,0)}, \chi^{(1)}\}$ and $\frac{\varepsilon^2}{2}\{\{f_2^{(0,0)}, \chi^{(1)}\}, \chi^{(1)}\}$ that belong to the classes $\mathcal{P}_{1,K}$ and $\mathcal{P}_{0,2K}$, respectively. Let us introduce $\hat{f}_1^{(1,1)} = \{f_2^{(0,0)}, \chi^{(1)}\}$ and $\hat{f}_0^{(1,2)} = \frac{1}{2}\{\{f_2^{(0,0)}, \chi^{(1)}\}, \chi^{(1)}\}$; it is easy to remark that

$$\langle \hat{f}_1^{(1,1)} \rangle = 0$$ meaning the average over the angles.

Thus, we can represent the expansion of the Hamiltonian as follows

$\hat{H}^{(1)} =$

$\lVert p \rVert^2$	$f_2^{(0,0)}$	0	0	0	...
$\lVert p \rVert$	$\underline{\omega}\cdot p$	$\varepsilon \hat{f}_1^{(1,1)}$	0	0	...
1	0	0	$\varepsilon^2 \hat{f}_0^{(1,2)}$	0	...
	$O(1)$	$O(\varepsilon)$	$O(\varepsilon^2)$	$O(\varepsilon^3)$	

Let us try again to remove the main perturbing term by a canonical transformation close to identity. Let us try with a generating function $\chi_2^{(1)} \in \mathcal{P}_{1,K}$ (so that $\{\underline{\omega}\cdot p, \chi_2^{(1)}\} \in \mathcal{P}_{1,K}$); therefore, we define

$$H^{(1)} = \exp\left(L_{\chi_2^{(1)}}\right) \hat{H}^{(1)}.$$

We determine $\chi_2^{(1)}$ so that

$$\left\{\underline{\omega}\cdot\underline{p}, \chi_2^{(1)}\right\} + \hat{f}_1^{(1,1)} = 0 ,$$

where $\hat{f}_1^{(1,1)}(\underline{p},\underline{q}) = \sum_{|\underline{j}|=1} \sum_{0<|\underline{k}|\le K} c_{\underline{j}\underline{k}}\, \underline{p}^{\underline{j}} \exp(i\underline{k}\cdot\underline{q})$. Let us write $\chi_2^{(1)}(\underline{p},\underline{q}) = \sum_{|\underline{j}|=1} \sum_{0<|\underline{k}|\le K} \varphi_{\underline{j}\underline{k}}\, \underline{p}^{\underline{j}} \exp(i\underline{k}\cdot\underline{q})$, therefore, we have

$$-i\underline{k}\cdot\underline{\omega}\, \varphi_{\underline{j}\underline{k}} + c_{\underline{j}\underline{k}} = 0 \quad \forall\, |\underline{j}| = 1 \text{ and } 0<|\underline{k}|\le K$$

$$\Rightarrow \varphi_{\underline{j}\underline{k}} = \frac{c_{\underline{j}\underline{k}}}{i(\underline{k}\cdot\underline{\omega})} \longrightarrow \text{small divisors.}$$

It is now easy to realize that three infinite sequences are generated at 0-degree by $\exp\left(\varepsilon L_{\chi_2^{(1)}}\right)\hat{f}_0^{(1,2)}$, at 1-degree by $\exp\left(\varepsilon L_{\chi_2^{(1)}}\right)\left(\underline{\omega}\cdot\underline{p} + \hat{f}_1^{(1,1)}\right)$ and 2-degree by $\exp\left(\varepsilon L_{\chi_2^{(1)}}\right)\hat{f}_2^{(0,0)}$.

This last remark encourages us to consider the general r-th step of normalization. Let us describe the expansion of $H^{(r-1)}$ as follows:

$H^{(r-1)} =$

$\Vert\underline{p}\Vert^2$	$f_2^{(r-1,0)}$	$\varepsilon f_2^{(r-1,1)}$	.....	$\varepsilon^{r-1} f_2^{(r-1,r-1)}$	$\varepsilon^{r} f_2^{(r-1,r)}$	...
$\Vert\underline{p}\Vert$	$\underline{\omega}\cdot\underline{p}$	0	...	0	$\varepsilon^{r} f_1^{(r-1,r)}$	...
1	0	0	...	0	$\varepsilon^{r} f_0^{(r-1,r)}$	...
	$O(1)$	$O(\varepsilon)$		$O(\varepsilon^{r-1})$	$O(\varepsilon^{r})$	

$\boxed{f_\ell^{(r-1,s)} \in \mathcal{P}_{\ell,sK}!}$

let us introduce $\hat{H}^{(r)} = \exp\left(\varepsilon^2 L_{\chi_1^{(r)}}\right) H^{(r-1)} = \exp\left(\varepsilon^2 L_{X+\xi\cdot q}\right) H^{(r-1)}$

where $\chi_1^{(r)}$ is determined so that we remove $f_0^{(r-1,r)}$ and $\langle f_1^{(r-1,r)} \rangle$. This is made by solving the equations

$$\varepsilon^2 \left\{ \underline{\omega}\cdot p, X^{(r)} \right\} + \varepsilon^r f_0^{(r-1,r)} = 0$$

and $\varepsilon^2 \left\{ f_2^{(0,r-1)}, \underline{\xi}^{(r)}\cdot q \right\} + \varepsilon^r \langle f_1^{(r-1,r)} \rangle = 0$.

This can be made by defining $X^{(r)} = \sum_{0<|\underline{k}|\le rK} x_{\underline{k}} \exp(i\underline{k}\cdot q)$

so that $x_{\underline{k}} = \frac{c_{\underline{k}}}{i(\underline{k}\cdot\underline{\omega})}$ ← small divisor, being $f_0^{(r-1,r)}(q) = \sum_{|\underline{k}|\le rK} c_{\underline{k}} \exp(i\underline{k}\cdot q)$.

Moreover we determine $\underline{\xi}^{(r)}$ so that

$C\underline{\xi} = \underline{\Delta\omega}$, being $\langle f_1^{(r-1,r)} \rangle = \underline{\Delta\omega}\cdot p$ and $f_2^{(0,0)} = \frac{1}{2} p\cdot Cp$.

It admits solutions provided $\det C \neq 0$.

We can describe the expansion of $\hat{H}^{(r)}$ in a similar way to what has been done for $H^{(r-1)}$ by putting a zero in the cell of the terms non-depending properly on the actions and of order ε^2. In order to define all the terms appearing in

$$\hat{H}^{(r)} = \underline{\omega}\cdot p + \sum_{0\le s\le r-1} \varepsilon^s f_2^{(r,s)} + \sum_{l=0}^{2} \sum_{s\ge r} \varepsilon^s \hat{f}_l^{(r,s)},$$

it is convenient to produce multiple ~~definitions~~ abuse of notations so to mimick a programming code. Let us start by putting

$$\hat{f}_\ell^{(r,s)} = f_\ell^{(r-1,s)} \quad \forall\, \ell = 0, 1, 2 \text{ and } s \geq 1$$
$$\text{or for } \ell = 2 \text{ and } s = 0 .$$

Therefore, we redefine the previous functions so that

$$\hat{f}_{\ell-j}^{(r,s+jr)} \hookleftarrow \frac{1}{j!} L^j_{\chi_1^{(r)}} f_\ell^{(r-1,s)} \quad \forall\, \ell = 1, 2,\ 1 \leq j \leq \ell,\ s \geq 1$$
$$\text{or for } s = 0,\ \ell = 2,\ 1 \leq j \leq 2$$

~~with~~ where the symbol $a \hookleftarrow b$ ~~deno~~ means that the quantity a is redefined so to put $\underset{\text{new value}}{a} = \overset{\text{old value}}{a} + b$.

The previous (re) definitions are such that (at the end of all the cycles) give the exact formulation of all the terms appearing in the expansion of $H^{(r)}$. Moreover, those (re) definitions respect the classifications in powers of ε and with respect to the class of functions $\mathcal{P}_{\ell,sK}$. In fact, it is easy to prove (by induction) that $\hat{f}_\ell^{(r,s)} \in \mathcal{P}_{\ell,sK}$.

Now, we introduce $\mathcal{H}^{(r)} = \exp\left(L_{\chi_2^{(r)}}\right)\hat{\mathcal{H}}^{(r-1)}$

where the generating function

$$\chi_2^{(r)}(p,q) = \sum_{|j|=1}\ \sum_{0<|k|\le rK} y_{jk}\, p^{j} \exp(i k\cdot q)$$

is determined so that

$$\left\{\omega\cdot p,\ \chi_2^{(r)}(p,q)\right\} + \hat{f}_1^{(r,r)} = 0$$

and, then, we have to put

$$y_{jk} = \frac{c_{jk}}{(i k\cdot\omega)} \longleftarrow \boxed{\text{small divisor}}$$

being the coefficients c_{jk} such that

$$\hat{f}_1^{(r,r)} = \sum_{|j|=1}\ \sum_{0<|k|\le rK} c_{jk}\, p^{j} \exp(i k\cdot q)\,.$$

At the end of the r-th normalization step we can write

$$\mathcal{H}^{(r)} = \omega\cdot p + f_2^{(0,0)}(p) + \sum_{l=0}^{2}\ \sum_{s>0} f_l^{(r,s)}(p,q)$$

where $f_l^{(r,s)}$ are initially set to be equal to $\hat{f}_l^{(r,s)}$ $\forall\ l=0,1,2$ and $s>0$ or for $l=2$ and $s=0$. Therefore, they are redefined many times so that

$$f_l^{(r,s+jr)} \hookleftarrow \frac{1}{j!} L^{j}_{\chi_2^{(r)}} \hat{f}_l^{(r,s)} \quad \forall\ l=0,1,2,\ s>0 \text{ and } j>0 \text{ or for } l=2 \text{ and } s=0$$

and $f_1^{(r,jr)} \leftarrow \frac{1}{j!} L^j_{\chi_2^{(r)}} \underline{\omega}\cdot\underline{p}$ $\forall j \geq 1$.

It is easy to check that $\longrightarrow f_\ell^{(r,s)} \in \mathcal{P}_{\ell,sK}$.

From a purely formal point of view, one expects that

$$\lim_{r \to +\infty} H^{(r)} = H^{(\infty)} := \underline{\omega}\cdot\underline{p} + \sum_{s\geq 0} \varepsilon^s f_2^{(\infty,s)}(\underline{p},\underline{q})$$

in Kolmogorov normal form.

A finite number of the algorithm steps (with suitably truncated Hamiltonians) are representable on a computer and, in practice, with the help of an algebraic manipulator, one could check that the norm of the generating functions and of all the terms grow geometrically with respect to s. This means that, if ε is small enough, the algorithm is really convergent.

However, a proof of the convergence of this algorithm based on classical type series expansions is very demanding and it will not be done in the present lectures.

We will prove the KAM theorem by using the quadratic ~~method~~ approach (that is a sort of generalization of the Newton method).

• Accumulation of the small divisors – a brief and informal discussion

Let us refer to the formal algorithm we have just described. Let us consider any norm somehow based on the sum of the (absolute values) of the coefficients. In order to try to simplify the situation let us avoid the distinction between $\chi_1^{(r)}$ and $\chi_2^{(r)}$. Let us try to describe ***naively*** the accumulation of the small divisors with the following scheme

order in ε \ norms	f	χ
1	1	$1\cdot\left(\frac{K}{\gamma}\right)^{\tau}$
2	$\chi^{(1)*}$	$\chi^{(1)}\cdot\left(\frac{2K}{\gamma}\right)^{\tau}$
3	$\chi^{(2)*}$	$\chi^{(2)}\cdot\left(\frac{3K}{\gamma}\right)^{\tau}$
$\vdots$	$\vdots$	$\vdots$
r	$\chi^{(r)}$	$\chi^{(r-1)}\cdot\left(\frac{rK}{\gamma}\right)^{\tau}$
$r+1$	$\chi^{(r)*}$	$\chi^{(r)}\cdot\left(\frac{(r+1)K}{\gamma}\right)^{\tau} = O\left(\varepsilon^{r+1}\cdot\left((r+1)!\right)^{\tau}\right)$

Such a scheme explains that there is not any hope to ~~have~~ obtain the convergence with a naive approach.

Indeed, the proof of the KAM theorem with the classical method requires to ~~give~~ prove strict rules on the accumulation

of the small divisors; those rules also depend on the degree in the actions.
Actually a small divisor introduced in a homological equation for $\chi_1^{(r)}$ can give contributions to the homological equation for $\chi_2^{(r)}$, $\chi_2^{(r+1)}$, ...
The key remark is that $\chi_2^{(r)}$ cannot give contributions to $\chi_2^{(s)}$ for any $s = r+1, \ldots, 2r-1$. The same holds true also for $\chi_1^{(r)}$.
~~The fact that~~ The accumulation of the most dangerous terms is such that a term appearing at order ε^r ~~is squared~~ give contributions at order ε^{2r}: this suggests that a quadratic scheme of convergence can be fruitful.

• Lie series – analytic settings

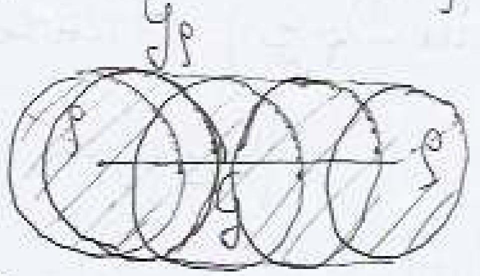
It is convenient to work with open complexified domains. Starting from any set $\mathcal{G} \in \mathbb{R}^n$, we consider sets of type

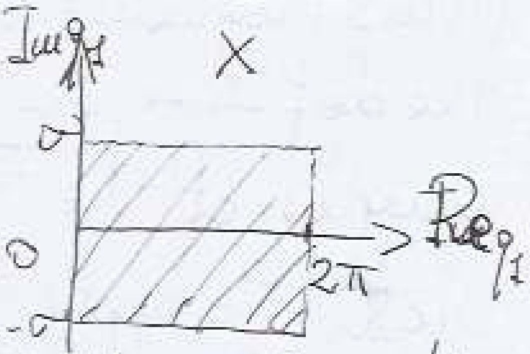
$$D_{\varrho,\sigma} = \mathcal{G}_{\varrho,\sigma} \times \mathbb{T}^n = \Big\{ (\underline{p}, \underline{q}) : \underline{p} \in \Delta_\varrho(\underline{p}') \text{ with } \underline{p}' \in \mathcal{G},\ \mathrm{Re}\,\underline{q} \in \mathbb{T}^n \wedge \mathrm{Im}\, q_j < \sigma \ \forall j = 1, \ldots, n \Big\};$$

where $\Delta_\rho(p')$ is a polydisk centered in p' and included in $\mathbb{C}^n$. This means that

$$\Delta_\rho(p') = \left\{ p \in \mathbb{C}^n : \left| p_j - p'_j \right| < \rho \ \ \forall j = 1, \ldots, n \right\}$$

In order to fix the ideas, it can be useful to refer to the figure on the right where the schematic situation (in the simplest case with $n=1$) is depicted.





A very comfortable way to ~~describe the~~ estimate the contributions of all the terms making part of the Lie series actually requires to work with the so called weighted Fourier norms. Let us introduce them. Let f be an analytic function on $\mathcal{D}_{\rho,\sigma'}$ with $\rho > 0$ and $\sigma' > \sigma > 0$, then the following norm of f on $\mathcal{D}_{\rho,\sigma}$ is well defined:

$$\| f \|_{\rho,\sigma} = \sum_{k \in \mathbb{Z}^n} \left| c_k \right|_\rho e^{|k|\sigma} , \quad \text{being } \left| c_k \right|_\rho = \sup_{p \in G_\rho} \left| c_k(p) \right| ,$$

where the Fourier expansion of f is

$$f(p, q) = \sum_{k \in \mathbb{Z}^n} c_k(p) \exp(i k \cdot q) .$$

let us assume that f is also bounded on $D_{\rho,\sigma'}$ (otherwise, ~~each~~ replace σ' with a value $\bar{\sigma}$ such that $\sigma' < \bar{\sigma} < \sigma$; then f is bounded in $D_{\rho,\bar{\sigma}}$), therefore

$$|c_k|_\rho \leq \sup_{(p,q)\in D_{\rho,\sigma'}} |f(p,q)| \cdot e^{-|k|\sigma'} .$$

(this inequality can be easily justified by recalling the definition, via the computation of the average over the angles, of the Fourier coefficients; then, it is just matter to apply the Cauchy's residue theorem)

The latter inequality ensures us that the decay of the Fourier coefficients is fast enough to make $\|f\|_{(\rho,\sigma)}$ convergent.

In order to deal with the proof of convergence of any perturbative scheme based on Lie series, the following estimates are really needed.

Lemma (6.14 in the Giorgilli's notes): Let f and g two analytic functions on $D_{\rho,\sigma}$ and $D_{(1-d')(\rho,\sigma)}$, respectively, with bounded weighted Fourier norms on their domain, being $0 \leq d' < 1$. Therefore, $\forall\, d \in (0, 1-d')$ the following inequalities hold true:

(A) $\left\|\frac{\partial f}{\partial p_j}\right\|_{(1-d)(\rho,\sigma)} \leq \frac{1}{d\rho}\|f\|_{\rho,\sigma}$ and

$\left\|\frac{\partial f}{\partial q_j}\right\|_{(1-d)(\rho,\sigma)} \leq \frac{1}{ed\sigma}\|f\|_{\rho,\sigma}$ $\forall j=1,\dots,n$;

(B) $\left\|\{f,g\}\right\|_{(1-d-d')(\rho,\sigma)} \leq \frac{2}{ed(d+d')\rho\sigma}\|f\|_{\rho,\sigma}\|g\|_{(1-d)(\rho,\sigma)}$

In order to prove the first inequality in (A), it is just matter to use the Cauchy inequality.
By exercise, let us prove the second inequality in (B).
First, let us write the Fourier expansion of

$$\frac{\partial f}{\partial q_j} = \sum_{\underline{k}\in\mathbb{Z}^n} i k_j\, c_{\underline{k}}(p)\exp(i\underline{k}\cdot q)\,.$$

$$\Rightarrow \left\|\frac{\partial f}{\partial q_j}\right\|_{(1-d)(\rho,\sigma)} \leq \sum_{\underline{k}\in\mathbb{Z}^n} |\underline{k}|\,|c_{\underline{k}}|_\rho\, e^{|\underline{k}|(1-d)\sigma}.$$

Since we have that $x^\alpha e^{-\beta x} \leq \left(\frac{\alpha}{\beta}\right)^\alpha e^{-\alpha} = \left(\frac{\alpha}{e\beta}\right)^\alpha$ $\forall x\geq 0$, $\alpha\geq 0$, $\beta>0$,

then $\left\|\frac{\partial f}{\partial q_j}\right\|_{(1-d)(\rho,\sigma)} \leq \frac{1}{ed\sigma}\sum_{|\underline{k}|\in\mathbb{Z}^n}|c_{\underline{k}}|_\rho\, e^{|\underline{k}|\sigma} = \frac{1}{ed\sigma}\|f\|_{\rho,\sigma}$.

By combining the techniques we used to prove the inequalities in (A), one can prove that in (B).

We are now ready to obtain also a result about multiple Poisson brackets.

Lemma (6.15 in Giorgilli's notes): let χ and f be analytic functions on $D_{\rho,\sigma}$ with finite weighted Fourier norms therefore

$$\left\| \frac{1}{s!} L_{\chi}^{s} f \right\|_{(1-d)(\rho,\sigma)} \leq \frac{1}{e^2} \left(\frac{2e}{d^2 \rho\sigma} \right)^s \|\chi\|_{\rho,\sigma}^{s} \|f\|_{\rho,\sigma}$$

$\forall\, s \geq 1$, $d \in (0,1)$.

Hint for the proof: Apply iteratively the inequality (B) of the previous lemma to $\frac{1}{j} \left\{ \frac{1}{(j-1)!} L_{\chi}^{j-1} f, \chi \right\}$ with $d' = \frac{(j-1)d}{s}$ and by replacing d with d/s, $\forall\, j = 1, \dots, s$.

Let us remark that in all these estimates we ~~can~~ take control of the growing coefficients generated by the derivatives at the price of making restrictions of the domains of analyticity. Too small restrictions of the domains allow just a weak control. On the other hand too strict control on the estimates is not allowed, because one could loose all the analyticity.